DICAL ROOM
RAL LIBRARY
V. OF MICH.

VOL. XXVII SEPTEMBER, 1936 NO. 6

OCT 19 1936

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

A Study of the Adjustment Difficulties of a Group of Women Teachers 401
LEIGH PECK

A Measure of the Singing and Rhythmic Development of Preschool Children 417
SERAPHINE S. SELTZER

An Unemphasized Factor in Current Theories Regarding Transfer of Training 425
ERNEST E. BAYLES

Note on a Graphic Method for Determining the Significance of the Difference between Group Frequencies 431
JOSEPH ZUBIN

Influence of Order of Presentation of Items in Ranking. 445
PERCIVAL M. SYMONDS

Relation of Handedness, Eye-sighting and Acuity Dominance to Reading. 450
ARTHUR I. GATES AND GUY L. BOND

A Formula for Intercorrelations among Multi-scores 457
ELMER B. ROYER

Over-compensation in Finger Timing as a Factor in Typewriting Errors 467
NEIL WARREN

Biserial r versus Pearson r as Measures of Test-item Validity . . 471
FRANCES SWINEFORD

Additional Data on Binocular Imbalance and Reading. 473
BRANT CLARK

Book Reviews 476

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15. 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Established 1910

EDITORS

JACK W. DUNLAP
Fordham University

HAROLD E. JONES
University of California

PERCIVAL M. SYMONDS
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Publishers.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVII | September, 1936 | Number 6

A STUDY OF THE ADJUSTMENT DIFFICULTIES OF A GROUP OF WOMEN TEACHERS

LEIGH PECK

The University of Texas

This study of the adjustment difficulties of a group of one hundred women teachers was made with the purpose: (1) To ascertain the proportion of the women teachers that were maladjusted, as judged by some objective standard, and to compare their adjustment with that of women students and with that of men; (2) to note the incidence of traits and habits that might be of especial significance in affecting a teacher's professional activities; (3) to study the relation to maladjustment of age, marital status, and extent and type of experience; (4) to determine what problems required from the teachers a conscious effort to adjust; (5) to note the relation of adjustment to intelligence; and (6) to compare self-estimates of adjustment with objective measurements of adjustment.

PREVIOUS STUDIES

Within the last ten years, several studies have been made of teachers' attitudes toward the adjustment problems of their pupils, and several recent textbooks in mental hygiene include chapters discussing the mental health of the teacher, assuming that teachers' own maladjustments influence their attitudes toward pupils' maladjustments. There have been, however, few investigations of the mental health of teachers.

In 1930, Goodwin Watson[1] reported a study of "Happiness among Adult Students of Education." His subjects were three hundred eighty-eight graduate students (averaging thirty years of age) in Teachers College, Columbia University. In securing his data,

[1] Watson, Goodwin: "Happiness among Adult Students of Education." *Journal of Educational Psychology*, Vol. XXI, 1930, pp. 79–109.

he had his subjects check their own estimates of their happiness on a five-point scale, had them choose phrases and adjectives descriptive of their emotional state, asked them to compare their present happiness with that in earlier years, gave them an Otis intelligence test, and secured answers to an extended questionnaire that included questions concerning childhood environment, scholarship, extra-curricular activities, recreations, and other topics. From these data, especially by comparing the fifty happiest with the fifty least happy, he derived an extensive list of conclusions, of which the following are most relevant to the present study: (1) Men believe themselves to be happier than women believe themselves to be; (2) age is not significant for happiness; (3) the married are happier than the unmarried; (4) intelligence bears no relation to happiness; (5) the general level of happiness can be measured with adequate reliability by a single check on one graphic scale.

Frances V. Mason,[1] in 1931, reported "A Study of Seven Hundred Maladjusted School Teachers." She secured her data from the records of seven hundred thirty-three patients (one hundred ninety men, five hundred forty-three women) in four hospitals for mental diseases, going back as far as 1880 for some of the records. She found that these teachers had been admitted to the hospitals at an earlier average age (57.9 per cent by the age of forty) than non-teachers. However, school troubles were mentioned as contributing causes of illness in only thirty-two of the two hundred fourteen cases for which information concerning contributing causes could be obtained. Paranoia was more often found among the teachers than among non-teachers. More than half of all the teachers were described as being, before their illness, sensitive, shy, or seclusive. As a group, they lacked social and recreational interests; reading, study, and religion were said to be their chief extra-curricular interests before their illness. About three-fifths were in poor physical condition.

PRESENT INVESTIGATION

The present investigation was concerned chiefly with the adjustment difficulties of a group of one hundred women teachers, enrolled in the author's classes in The University of Texas during the summer of 1935. As control groups, fifty-two women students who were not teachers, and twenty-six men (teachers and prospective teachers),

[1] Mason, Frances V.: "A Study of Seven Hundred Maladjusted School Teachers." *Mental Hygiene*, Vol. XV, 1931, pp. 576–599.

in the same classes, were also studied. These one hundred seventy-eight subjects checked the Thurstone Personality Schedule, as a measure of adjustment or maladjustment; took the Otis Self-Administering Test of Mental Ability, Higher Examination, Form A; filled out a personal data sheet, giving age, marital status, teaching experience, if any, and estimate of own adjustment, on a five-point scale; and furnished an informal written statement of the problems to which the subject was making a conscious effort to adjust. In supplying this information, the subjects remained entirely anonymous. At the beginning of the investigation, they drew for slips of paper on which numbers had been written, and they signed all papers pertaining to the investigation by number rather than by name.

RESULTS

1. *Scores on the Thurstone Personality Schedule*

The highest possible score on the Thurstone Personality Schedule is two hundred twenty-three; scores between thirty and fifty-nine are considered as average. The mean scores of all three groups included in the present study fall within the average limits; however, the women teachers evidence a greater number of maladjustments than do the women students or the men; and the women as a whole evidence a greater number of maladjustments than do the men. For the men, the mean is forty-one, the range two to one hundred; for the women students, the mean is forty-five, the range ten to ninety-eight; for the women teachers, the mean is fifty, the range eight to one hundred eighty-five. These figures are in agreement with Watson's finding that men believe themselves to be happier than women believe themselves to be.

Only five of the women students stated that they did not intend to teach; their mean score is forty-nine, as compared with the mean of forty-five for the entire group of women students. Thus, as far as can be judged from this small group, there is no evidence that it is the women students already maladjusted who tend to enter the teaching profession.

No difference in mean score is observed in comparing the men students (all prospective teachers) with the men teachers; the mean score of the former ($N = 20$) is 41.2, for the latter ($N = 6$), 41.5. Thus, though the women teachers are more maladjusted than the women students, the men teachers are as well adjusted as the men

students. Apparently, if teaching as a profession tends to produce maladjustments, the occupational hazards are more serious for women than for men.

Scores on the Thurstone Personality Schedule may be interpreted in terms of letter grades *A*, *B*, *C*, *D*, and *E*, corresponding to the descriptive phrases "extremely well adjusted," "well-adjusted," "average," "emotionally maladjusted," and "should have psychiatric advice." These letter grades and descriptive phrases were arbitrarily selected by Thurstone to designate five groups of scores: 0–14, *A*; 15–29, *B*; 30–59, *C*; 60–79, *D*; and 80 or above, *E*.

TABLE I.—COMPARISON OF SCORES ON THURSTONE PERSONALITY SCHEDULE

Score	Grade	Description	Chicago freshmen (N = 682)	Percentages making score		
				Women teachers (N = 100)	Women students (N = 52)	Men (N = 26)
0–14	*A*	Extremely well adjusted.	16	05	06	11.5
15–29	*B*	Well adjusted.	29	14	21	15
30–59	*C*	Average.	41	48	50	58
60–79	*D*	Emotionally maladjusted.	10	21	08	11.5
80–	*E*	Should have psychiatric advice.	04	12	15	04

Table I shows the classification in terms of letter grades of the subjects of the present investigation, and includes a comparison of their grades with those of subjects used by Thurstone in standardizing the test. It will be seen from Table I that under a system of classification so liberal that Thurstone could designate forty-five per cent of his subjects as "well adjusted" or "extremely well adjusted," forty-one per cent as "average," and only fourteen per cent as "maladjusted" or "in need of psychiatric advice," thirty-three per cent of the women teachers are classified as "maladjusted" and twelve per cent of these are "in need of psychiatrice advice." Only nineteen per cent are "well adjusted" or "extremely well adjusted." It may also be seen from Table I that the women students, though making a better showing than the women teachers, include more "maladjusted" and fewer "well adjusted" individuals than do the members of the Thurstone group. The men, though showing a percentage of maladjustment almost as low as that in Thurstone's group, have a relatively

low percentage of the well adjusted. In other words, even the men teachers and prospective teachers included in the present study are not, on the whole, quite so well adjusted as the subjects upon whom Thurstone's test was standardized; the women students are still less well adjusted; and the women teachers least of all, to the extent that thirty-three per cent are classified as definitely maladjusted and twelve of these in need of psychiatric advice.

2. *Comparison of Responses to Selected Questions*

From the two hundred twenty-three questions in the Thurstone Personality Schedule, forty-four items were selected arbitrarily by the present investigator as relating to traits or habits that might be of especial significance in affecting a teacher's professional activities. These forty-four items are shown in Table II, classified under the headings: (1) Undesirable emotional reactions, (2) unfavorable personality traits, (3) social handicaps, (4) poor work habits, (5) incapacitating physical symptoms, (6) history of previous maladjustments, and (7) sex problems. The percentage of teachers answering affirmatively each of the forty-four questions is indicated in the first column in Table II. It is interesting to speculate as to the possible effect on pupils of the fact that nearly one-third of these teachers feel themselves deserving of a better lot in life, one-fifth report themselves to be frequently in low spirits, and one-fourth often in a state of excitement; one-third describe themselves as nervous, one-sixth admit that they lose their tempers quickly, one-fourth are upset easily; one-third are shy, one-fourth are critical of other people; one-fourth do not plan their work ahead, and an equal number dislike to take responsibility; one-fourth frequently suffer from indigestion, and one-fifth are tired most of the time; one-tenth have had a nervous breakdown, and one-fourth fear insanity; one-fourth are suffering from a conflict between sex and morality.

The second column in Table II indicates the number of women students answering affirmatively the forty-four selected questions, and the third column compares the teachers' responses with the students' responses by giving the critical ratio (that is, the difference divided by the sigma of the difference). If a difference equal to at least three sigma be used as the criterion of a significant difference, it will be seen that the responses of the women teachers differ from those of the women students significantly only in item 22: "Saying things that hurt people's feelings." Here the affirmative responses

TABLE II.—COMPARISON OF RESPONSES TO THURSTONE PERSONALITY SCHEDULE

Responses	Percentages			Percentages		
	Women teachers (N = 100)	Women students (N = 52)	Critical ratio	Women (N = 152)	Men (N = 26)	Critical ratio
I. Undesirable emotional reactions.						
1. Feels unadjusted to life.	17	29	1.64	21	37	1.59
2. Feels life a great burden	08	06	.47	07	11	.61
3. Feels deserving of a better lot in life	30	21	1.39	27	26	.10
4. Worries over possible misfortunes	37	33	.49	36	26	1.05
5. Worries over the tasks of the day	06	10	.83	07	07	.00
6. Moody	37	29	1.27	34	05	5.05
7. Frequently in low spirits	22	15	1.08	20	18	.24
8. Often in a state of excitement	25	21	.56	23	00	6.74
II. Unfavorable personality traits.						
9. Nervous	33	33	.00	33	18	1.77
10. Lacks self-confidence	30	30	.00	30	30	.00
11. Feelings easily hurt	37	40	.36	38	52	1.32
12. Disturbed by criticism	31	23	1.07	28	15	1.65
13. Loses temper quickly	16	08	1.52	13	15	.26
14. Confused easily	30	35	.62	32	11	2.91
15. Discouraged easily	45	50	.58	47	30	1.72
16. Upset easily	23	21	.28	22	00	6.54
III. Social handicaps.						
17. Shy	33	35	.24	34	33	.10
18. Has difficulty in speaking in public	46	44	.23	45	22	2.53
19. Has difficulty making friends	16	23	1.01	18	15	.39
20. Finds books more interesting than people	18	18	.00	18	30	1.26
21. Seldom interested in people	08	06	.47	07	15	1.09
22. Says things that hurt people's feelings	15	02	3.20	10	11	.15
23. Has habit of contradicting people	10	12	.37	11	11	.00
24. Critical of other people	24	13	1.74	20	26	.65
IV. Poor work habits.						
25. Mind wanders badly	16	12	.69	14	15	.13
26. Does not plan work ahead	23	21	.28	22	22	.00
27. Leaves many tasks unfinished	28	23	.65	26	18	.96
28. Dislikes to take responsibility	26	23	.40	25	18	.84
V. Incapacitating physical symptoms.						
29. Poor appetite	07	04	.80	06	04	.44
30. Frequent indigestion	26	18	1.16	23	22	.11
31. Poor sleep	09	10	.19	09	11	.30
32. Many headaches	10	10	.00	10	11	.15
33. Fatigue upon awaking	21	19	.29	20	22	.30
34. Tires of work quickly	27	36	1.12	30	18	1.43
35. Feels tired most of time	21	30	1.19	24	22	.30
VI. History of maladjustments.						
36. Unhappy home environment	06	04	.55	05	07	.37
37. Unhappy childhood	15	18	.46	16	04	2.47
38. Unhappy adolescence	20	12	1.32	17	07	1.70
39. Nervous breakdown	11	08	.58	10	04	1.32
40. Fear of insanity	24	15	1.38	21	11	1.43
41. Desire to commit suicide	14	18	.62	15	07	1.38
VII. Sex problems.						
42. Ignorance of sex	14	15	.16	14	00	5.00
43. Feeling of sexual inferiority	15	10	.91	13	07	1.05
44. Conflict between sex and morality	25	21	.56	24	37	1.28

of the teachers exceed those of the students so appreciably that the critical ratio is 3.2.

The fourth column in Table III presents the percentage of affirmative responses made by the women teachers and women students combined, and the fifth column shows the percentage of affirmative responses made by the men. The sixth and last column compares the responses of the women with those of the men, giving the critical ratio (the difference divided by the sigma of the difference). If a difference of at least three sigma be again taken as the criterion of a significant difference, we find that the women differ from the men with respect to traits that are possibly related to differences in physical constitution: The women (1) are more moody; (2) are more frequently in a state of excitement; and (3) are upset more easily. The women also more often consider themselves ignorant of sex. The differences closely approach the criterion of three sigma with respect to three other items: The women (1) are confused more easily; (2) have more difficulty in speaking in public; and (3) have had a more unhappy childhood than the men.

In the remaining thirty-seven of the forty-four items, the differences between the responses of the men and those of the women do not satisfy the criterion for significant difference.

3. *Factors Apparently Related to Personality Scores*

Teachers are often told that if they can effect a happy adjustment of the children in their care, they will promote the future mental health of these children, grown to adulthood. This theory is somewhat substantiated by the records of the subjects of the present study. In Table III are tabulated separately the scores on the Thurstone Personality Schedule of the subjects reporting a happy childhood and adolescence and the scores of those reporting unhappiness in childhood or adolescence or both. In each group, a large difference in score favors the subjects who had a happy childhood and adolescence: among the women teachers, an average of forty-four symptoms of maladjustment for "happy" group, as compared to sixty-two for the "unhappy"; among the women students, an average score of thirty-nine, as compared to sixty-four; and among the men, an average score of 37.5, as compared to fifty.

Watson, in his study of "The Happiness of Adult Students of Education," found no indication that age affected the scores of his subjects. Age does, however, affect markedly the scores on the Thur-

stone Personality Schedule of the subjects of the present investigation. As shown in Table III, the scores of the men decrease (improve) steadily with age: twenty years and under, average score sixty-three; twenty-one to twenty-five years, average score forty; and twenty-six to thirty years, average score twenty-two. There is also a slight decrease with age in the scores of the women students: twenty years and under, average score forty-seven; twenty-one to twenty-five years, average score forty-five. In the case of the women teachers, the number of symptoms of maladjustment is highest in the years twenty-six to thirty, an average of fifty-six, and drops to an average of forty-one for the years of thirty-one and above. The teachers' own reports of their adjustment problems indicate that in the late twenties difficult decisions have to be made with respect to marrying or staying single, locating a relatively permanent position, and completing one's

TABLE III.—FACTORS RELATED TO AVERAGE SCORES ON THURSTONE PERSONALITY SCHEDULE

	Women teachers (N = 100)		Women students (N = 52)		Men (N = 26)	
	Number reporting	Average score	Number reporting	Average score	Number reporting	Average score
1. Happiness in childhood and adolescence.						
(1) Group reporting happiness	70	44	42	39	23	37.5
(2) Group reporting unhappiness	30	62	10	64	3	50
2. Age.[1]						
(1) Twenty and under	8	53	35	47	5	63
(2) Twenty-one to twenty-five	46	51	13	45	15	40
(3) Twenty-six to thirty	21	56	..	....	5	22
(4) Thirty-one and over	19	41				
3. Marital status.[1]						
(1) Single	68	49.5	43	47	20	45
(2) Married	16	53	6	43.5	6	41
(3) Widowed	9	35				
4. Number of years of teaching experience.[1]						
(1) One year	10	59				
(2) Two to five years	47	51				
(3) Six to ten years	19	47				
(4) Over ten years	15	42				
5. School grades taught.[1]						
(1) Rural	20	47				
(2) Primary	50	53				
(3) Intermediate	14	47				
(4) High School	10	40				

[1] This information was not furnished by some subjects.

education. The necessity for these decisions does not become so pressing in earlier years; and in later years the decisions have already been made and the individual reconciled to her lot. Thus the late twenties stand out as a relatively perturbed period in lives of many women teachers.

Watson found that the married were happier than the unmarried. The figures given in Table III indicate that, among the subjects of the present investigation, the married women students are somewhat better adjusted than the unmarried (with an average score on the Thurstone Personality Schedule of 43.5 as compared to forty-seven), and that the married men are somewhat better adjusted than the single men (with an average score of forty-one as compared to forty-five). Among the women teachers, however, the married women are worst adjusted, with an average score of fifty-three; the single women next, with an average score of 49.5; and the widows best adjusted, with an average score of thirty-five. The favorable score of the widows, as compared with the single women, may be influenced by an age factor, for the average age of the widows is 37.5, of the single women, twenty-four; and, as already noted, there is a decrease in the average number of maladjustment symptoms after the age of thirty is reached. However, the difference between the married women and the widows cannot so easily be explained as the result of an age factor, for the average of the married women is thirty-two, within the age limit in which symptoms of maladjustment tend, in general, to decrease. These married women teachers may be a group selected (that is, forced to teach in the winter and attend school in the summer) by some factor strongly tending in itself to produce maladjustment. That is, many of them might not be professionally active unless some financial exigency or family difficulty forced them out of their homes.

Teachers often allege that "the first year is the hardest." This assertion is substantiated in part by the fact, as shown in Table III, that the women teachers with only one year of experience record on the average fifty-nine symptoms of maladjustment on the Thurstone Personality Schedule; the teachers with two to five years of experience, fifty-one; those with six to ten years of experience, forty-seven; and those with more than ten years of experience, only forty-two.

Table III further shows that the highest average number of maladjustment symptoms is recorded for the primary teachers, with an average score of fifty-three on the Thurstone Personality Schedule; the rural teachers and the intermediate teachers average

forty-seven; and the high-school teachers manifest fewest symptoms of maladjustment, with an average score of forty.

A further analysis of the factors apparently related to the personality scores and grades of the women teachers is presented in Table IV. An examination of Table IV shows that the women teachers making personality grades of *B* and *A* ("well adjusted" and "extremely well adjusted") are distinguished from those making a grade of *C* (the "average" group) by being older (thirty-one or thirty-two years as compared to twenty-six years) and by having taught longer (nine years as compared to five years). Moreover, the "extremely well adjusted" teachers report an average of only two problems to which they are consciously endeavoring to adjust, as compared to 4.2 problems reported by the "well adjusted" group and 4.8 problems reported by the "average" group. The *D* and *E* groups ("emotionally maladjusted" or "in need of psychiatric advice") are slightly younger than the "average" group, but probably not significantly so (25.9 and 24.8 years as compared with twenty-six years), and have taught approximately the same length of time as the "average" group (5.2 and 5.1 as compared to 5.3 years). They do, however, differ noticeably from the "average" group in the number of problems to which, according to their own report, they are consciously trying to adjust (6.75 and six, as compared to 4.8 for the "average" group).

TABLE IV.—FACTORS RELATED TO THE PERSONALITY GRADES OF THE WOMEN TEACHERS

Grade	Percentage making grade (N = 100)	Average age of group	Average teaching experience (in years)	Average number of adjustment problems listed	Average Otis score
A	5	31.0	9.4	2.0	52.6
B	14	32.5	9.0	4.2	52.0
C	48	26.0	5.3	4.8	50.6
D	21	25.9	5.2	6.75	54.8
E	12	24.8	5.1	6.0	51.7

4. *Sources of Maladjustment Reported by Women Teachers*

The subjects of the present study were asked to write out, in an informal fashion, a statement of the problems to which they were

consciously trying to adjust, the problems which they felt to be more or less a source of maladjustment. Some responded in the briefest possible phrases; others submitted as many as eight or ten closely written pages detailing the difficulties which confronted them. The problems reported by the women teachers are summarized in Table V. These one hundred subjects list a total of five hundred four problems, an average of five each. As already noted in Table IV, however, the "extremely well adjusted" teachers list an average of only two problems each, while the "maladjusted" teachers list on the average of six or more problems each. As indicated in Table V, one-fifth of all the women teachers report difficulty in securing comfortable living accommodations, one-third deplore the lack of congenial associates, one-fourth suffer from lack of recreational facilities, one-fifth find their social activities too narrowly restricted by community prejudices, and one-fifth consider themselves overburdened with religious and educational activities outside the school. These facts deserve emphasis in view of Mason's finding, that lack of recreational and social interests seemed related to the mental illnesses of the teachers whose records she studied.

One-half the women teachers with whom the present study is concerned complain of heavy teaching load. Numerous studies have been made of the effect of large classes on efficiency in academic instruction. Perhaps some attention should be directed to the effect on the mental health of the teacher of the continued strain of adjusting her own personality to the personalities of large groups of children. One-fifth of these teachers fear the loss of their positions. One-fourth consider their salaries inadequate, and one-fifth have financial problems of a personal nature, aside from any question as to the adequacy of their salaries. One-fifth report health problems serious enough to be sources of maladjustment; overfatigue is most often mentioned.

5. *Intelligence Test Scores*

The scores on the Otis Self-Administering Test of Mental Ability (Higher Examination) may be interpreted in terms of point scores, letter grades, or intelligence quotients. The average point score for the women teachers is fifty-two, with a range of twenty-seven to seventy-three; for the women students fifty-three, with a range of thirty-seven to seventy-three; for the men fifty-seven, with a range of forty-one to sixty-nine. In each group the average letter score is

TABLE V.—SOURCES OF MALADJUSTMENT LISTED BY ONE HUNDRED WOMEN TEACHERS

I. Community conditions.		
1. Adjustment to new community		10
2. Lack of desirable living arrangements		21
3. Lack of congenial associates		33
4. Lack of recreational facilities		24
5. Restrictions on teachers' social activities and personal conduct		19
6. Imposition of extra duties (such as Sunday School teaching or P.T.A. work)		21
II. Teaching conditions.		
1. Friction with pupils		4
2. Work with backward or maladjusted children		11
3. Friction with parents		5
4. Friction with other teachers		5
5. Friction with superintendent, principal, or supervisor		13
6. Friction in community affecting the school		11
7. Heavy teaching load (too many classes, too large a group of pupils)		49
8. Lack of equipment		13
9. Burdensome clerical work		5
10. Heavy extra-curricular duties		11
11. Uncongenial tasks (assignment to teach in undesirable part of town, to teach subjects not liked, or to teach grades not preferred)		8
12. Burdensome requirements for professional self-improvement		5
13. Uncertainty of tenure in position		19
14. Inadequate salary		24
15. Delay in payment of salary		5
III. Personal problems.		
1. Financial problems		20
2. Health problems		20
3. Worry over welfare of relative or friend		14
4. Inharmonious relationships with members of family or other intimates		7
5. Frustration of ambitions		15
6. Mental and emotional conflicts		15
7. Personality problems		97
(1) Shyness, seclusiveness	18	
(2) Over-conscientiousness	12	
(3) Tendency to worry	17	
(4) Inferiority feelings	39	
(5) Miscellaneous	11	
Total		504

B, with a range from *C*− to *A* for the women teachers, *C* to *A* for the women students, and *C*+ to *A* for the men. The Otis IQ's for the women teachers average one hundred ten, with a range of eighty-five to one hundred thirty-one; for the women students, one hundred

eleven, with a range of ninety-five to one hundred thirty-one; for the men, one hundred fifteen, with a range of ninety-nine to one hundred twenty-seven. The Otis test was, of course, inadequate to measure the ability of the more intelligent subjects.

The correlations between intelligence, as measured by the Otis test, and adjustment, as measured by the Thurstone Personality Schedule, are positive but very low: For the women teachers, $+.06 \pm .06$; for the women students, $+.03 \pm .08$, and for the men, $+.14 \pm .12$. Evidently intelligence is but a negligible factor in adjustment, and possibly of even less value to the women than to the men. It will be recalled that Watson found no relationship between happiness and intelligence.

6. *Correlations of Personality Scores with Self-estimates of Adjustment*

On their personal data sheets the subjects of the present investigation recorded self-estimates of their adjustment on a five-point scale: Unusually good, above average, average, below average, or unusually poor. The correlations are positive but low between these self-estimates of adjustment and the letter grades indicative of degrees of adjustment (*A*, *B*, *C*, *D*, *E*) derived from the scores on the Thurstone Personality Schedule. For the women teachers the correlation is $+.10 \pm .06$; for the women students, $+.38 \pm .09$; for the men, $+.31 \pm .12$. If, then, scores on the personality test can be taken as criteria of adjustment, the subjects of the present study are quite unsuccessful in evaluating their own mental health, and the women teachers are even less accurate in their estimates then are the women students and the men.

Watson referred to no objective criterion when he said that "the general level of happiness can be measured with adequate reliability by a single check on one graphic scale." He simply compared the subjects' self-estimates on a five-point scale with other types of self-estimates (such as the checking of descriptive phrases), finding a high correlation between the two.

SUMMARY AND CONCLUSIONS

The present investigation sought to ascertain the adjustment difficulties of a group of one hundred women teachers. Control groups, consisting of fifty-two women students (nearly all prospective teachers) and twenty-six men (teachers and prospective teachers) were also studied. Data were secured by means of the Thurstone

Personality Schedule, the Otis Test of Mental Ability, personal data sheets, and informal statements of the problems to which the subjects were making a conscious effort to adjust. The following conclusions seem justified:

1. The women teachers are not so well adjusted as the women students, and the women as a whole are not so well adjusted as the men, if the Thurstone Personality Schedule can be considered a valid measure of maladjustment.

2. One-third of the women teachers are definitely maladjusted, and one-sixth need psychiatric advice, as judged by the Thurstone Personality Schedule. Only one-fifth can be classified as well-adjusted.

3. The symptoms of maladjustment most common among the women teachers include the following: One-third feel deserving of a better lot in life, one-fifth are frequently in low spirits, one-fourth are frequently in a state of excitement; one-third describe themselves as nervous, one-sixth admit that they lose their tempers quickly, one-fourth are upset easily; one-third are shy; one-fourth are critical of other people; one-fourth do not plan their work ahead, and an equal number dislike to take responsibility; one-fourth frequently suffer from indigestion, and one-fifth are tired most of the time; one-tenth have had a nervous breakdown, and one-fourth fear insanity; one-fourth are suffering from a conflict between sex and morality.

4. Though the women teachers report a greater number of maladjustment symptoms than do the women students, no appreciable difference in the percentage reporting any one given symptom is found except with respect to the item, "saying things that hurt people's feelings." This is reported significantly more often by the teachers than by the students.

5. Though the women as a whole report a greater number of maladjustment symptoms than do the men, reliable differences are found with respect to the following items only: The women are more moody, are more frequently in a state of excitement, and are upset more easily; they also more often consider themselves ignorant of sex. The criterion for a reliable difference is approached in three other items: The women are confused more easily, have more difficulty in speaking in public, and have had a more unhappy childhood than the men.

6. More symptoms of maladjustment are reported by subjects who say they were unhappy in childhood or adolescence than by those who say they were happy then.

7. In general, adjustment improves with age. The peak of maladjustment, however, is reached by the women teachers between the ages of twenty-six and thirty.

8. Among the men and among the women ~~teachers~~ students?, the married are better adjusted than the unmarried; but among the women teachers studied, the widows are best adjusted, the single women next, and the married women worst adjusted.

9. Among the women teachers, adjustment improves as teaching experience increases; those who have taught one-year are worst adjusted, those who have taught more than ten years are best adjusted.

10. Of the women teachers, those in the primary grades are worst adjusted, those in the high school best adjusted.

11. Relatively few problems (an average of two) causing a conscious effort to adjust are reported by the group of teachers classified as "extremely well adjusted," but a relatively large number (an average of six or more) is reported by the group classified as "maladjusted."

12. The problems most often reported by the women teachers include the following: One-fifth find difficulty in securing comfortable living accommodations, one-third lack congenial associates, one-fourth lack recreational facilities, one-fifth find their social activities restricted by community prejudices, and one-fifth are overburdened by religious and educational activities outside the school; one-half consider their teaching load too heavy; one-fifth fear the loss of their positions, one-fourth consider their salaries inadequate, and one-fifth have financial problems of a personal nature, aside from any question as to the adequacy of their salaries; one-fifth report health problems serious enough to be sources of maladjustment.

13. The intelligence test scores of the men (teachers and prospective teachers) are slightly higher than those of the women (teachers or students).

14. The correlations between intelligence scores and adjustment scores are so low as to be negligible.

15. The correlations between the self-estimates of adjustment and the adjustment test scores are so low as to indicate that the subjects are unable to judge their own mental health.

SUGGESTIONS FOR FURTHER STUDY

The investigation reported has left untouched the question as to how the teachers' own adjustment difficulties affect their attitude toward their pupils' problems. The present author hopes to study

that question, using as subjects a group of teachers attending summer school, and utilizing techniques already developed to measure (*a*) the teachers' own adjustment and (*b*) their attitude toward the adjustment problems of pupils. A more critical study would necessitate the investigation during the school year of the mental health of a group of teachers in service and the coincident investigation of the mental health of the pupils being taught by these teachers.

A MEASURE OF THE SINGING AND RHYTHMIC DEVELOPMENT OF PRESCHOOL CHILDREN[1]

SERAPHINE S. SELTZER

The Merrill-Palmer School

The purpose of this study was to devise a measure, simple to understand and use, with which the musical development of preschool children, specifically, their singing and rhythmic development, could be evaluated.

Seashore's measures of musical talent[2] and the Kwalwasser-Ruch tests of musical accomplishment[3] are too involved and difficult for use with young children. Various studies[4] of the musical ability of young

---

[1] Adapted from a study made under the direction of Dr. Rachel Stutsman of the Merrill-Palmer School, and accepted by the College of Education, Wayne University, in partial fulfillment of the requirements for the Master's degree. The author wishes to thank Dr. Stutsman, Miss Leone Chesire, Miss Louise Conklin, Miss Jessie Wedin, and Dr. Walter Bergman, all of whom helped in various ways in the study.

[2] Seashore, C. E.: *The Psychology of Musical Talent.* Silver, Burdett & Co., 1919.

[3] Kwalwasser-Ruch: *Test of Musical Accomplishment.* Extension Division, University of Iowa, 1924. Information tests by which children are graded on their musical information.

[4] Vance, T. F. and Grandprey, M. B.: "Objective Methods of Ranking Nursery School Children on Certain Aspects of Musical Capacity." *Journal of Educational Psychology*, Vol. XXII, November, 1931, pp. 577–585. Ratings of children on seven items, including response to music under different circumstances, beating time to phonograph records, imitating teachers in beating rhythmical patterns on the triangle, and musical environments in the home.

McGinnis, Esther: "Seashore's Measures of Musical Ability Applied to Children of the Preschool Age." *American Journal of Psychology*, Vol. XL, 1928, pp. 620–623. In this study three of the Seashore tests, of pitch, intensity, and consonance, were selected for experiment with preschool children. The results showed that these tests cannot be used with young children in their present form, but with suitable modification might become valuable research instruments for use with young children.

Williams, H. M., Sievers, C. H., Hattwick, M. S.: *Measurement of Musical Development.* University of Iowa Studies in Child Welfare, Vol. VII, No. 1, 1933. A series of studies of which the first presents data on the rhythmic performances and vocal control of pitch of preschool children; the second, on an attempt to measure the ability of the individual child to tap in response to patterns differing in speed and complexity; the third, the pitch inflection in the speech of preschool children. These monographs are technical and concerned with the techniques of devising scales of measurement, not with educational procedures.

children have been made, but none offers a quantitative method of evaluating the musical development of young children which will show the musical needs and record the rate of musical development of the individual child. Certainly the observant teacher of music can discover great differences in the musical abilities of his young pupils and recognize their needs fairly well without a quantitative method of evaluation. However, the teacher of many children soon forgets the earlier musical status of the children and finds it difficult to say what progress they have made. If records are to be kept, his judgment of the ability and progress of the children, in the absence of a rating device, is not easily transferred to paper. Moreover, a quantitative method of rating the children's development in music enables the teacher to determine which of two methods of instruction will bring about the greater rate of progress in the total level of the group.

Some may object to an attempt to measure ability and progress in an art, such as music. However, at the Merrill-Palmer School, where the study was made, many useful quantitative measures of the development of young children have been devised, some of them dealing with factors which appear just as elusive as the artistic skill of singing and rhythmic response. The present study is an attempt to add one more device to the means of measuring the different aspects of artistic development, and to make more nearly complete the means of recording the entire development of the child. To the music teacher of young children, in particular, the scale is intended to offer a means of determining what teaching procedures are most successful, what the child's musical status is, and what progress he makes under instruction.

SPECIAL REQUIREMENTS OF A SCALE MEASURING MUSICAL DEVELOPMENT

In general, tests evaluating the abilities of young children measure their performance in controlled situations. Musical development might be measured in this way. However, such a measurement requires standards of achievement, and without elaborate procedures would be subjective, offering no more real quantitative information than a rating, despite a complex set-up. Moreover, in such a situation the group influence, the spontaneity of the child, and his lack of self-consciousness when expressing himself in a less controlled situation would be lost. It is desirable also that the child's response to music be tested in many situations, not a single one, before a decision is made concerning his ability or attainment.

The most suitable way of studying the child's musical development seems to be to observe his progress from day to day and to record it by means of a rating device defining the type of observation in concrete situations.

DEVISING THE SCALES

Since no earlier scale existed as a pattern to follow or deviate from, it was necessary to experiment with methods devised to evaluate other aspects of development. The method of evaluating aspects of personality in young children, devised at the School, appeared to lend itself to the measurement of musical development. This method is an adaptation of that described by Thurstone in his *Measurement of Attitude*.[1] Thurstone explains the method as an attempt to represent the distribution of a group on a specific issue in the form of a frequency distribution, represented by a linear continuum, extending from extreme lack of development, or absence of the attitude being measured, to extreme development. In the present scale, the positions in the continuum of particular statements concerning singing development or rhythmic response were determined by qualified judges. The progress of the preschool child to be measured is determined by checking the statements descriptive of his degree of development, and the mean score of the statements checked represents his general musical development.

The most exacting task in formulating the scale was the setting up of two lists of statements, one dealing with singing development, another with rhythmic ability. These were formulated on the basis of the music program with the Merrill-Palmer nursery-school children and from other available methods and materials used in the musical education of young children. These statements were evaluated and revised, and many of them discarded. The judges were asked to number the statements from one to eleven, on the basis of whether they represented, at one, the greatest possible lack of progress in singing or rhythmic development, or, at eleven, the greatest possible development, or some point between, with the midpoint at six. In singing development, statements concerned not only the child's singing, but also his response to music sung to him and appreciation of music during the singing period in the nursery school. In rhythmic development, the state-

---

[1] Thurstone, L. L. and Chave, E. J.: *The Measurement of Attitude.* University of Chicago Press, 1929.

ments concerned rhythmic activities expressed through the child's body as well as performance on percussion band instruments.

The original lists of statements were evaluated by a total of ninety-seven judges, in four groups. Ideally all these judges should have been persons experienced in music at the preschool level, but, since it was not possible to secure sufficient judges with these qualifications, the nearest approximation was made in securing the assistance of three groups of teachers and students associated with Detroit Teachers College. Some of these were students with special training in music. The greater number of the judges had had experience in teaching music in the grades and were at the time registered in music courses. The fourth group of judges, from the Merrill-Palmer School, comprised five women who had had experience in music for young children.

The returned statements were tabulated to show the opinion of each judge on each statement, the cumulative frequency of each statement was determined, and a percentile ogive curve drawn on the basis of the results. The cumulative frequencies of each statement were plotted on a graph in order to determine both the scale and ambiguity values of the statement. The scale value of each statement is the point where the curve passes the fifty per cent level, *i.e.*, the median, and the ambiguity value is determined by finding the difference between the values of the points where the curve crosses the seventy-fifth and twenty-fifth per cent levels. The final statements were selected from the original list by the criteria of least ambiguity, greatest relevance, clear-cut scale value, failure to duplicate other statements, and, in case of difficulty, by a final checking of the ratings with those of the five judges who had experience of music for preschool children. On the basis of these criteria the final scales, shown in the following pages, were arrived at. The final rating scale of singing development includes forty-two statements, selected from an original list of ninety-eight, and the final scale of rhythmic development forty-four statements, selected from an original ninety-four.[1]

The number in parenthesis before each statement is the scale value of the statement. This rating key is placed beside the statements only after they have been checked for the individual child. Only those statements true of the child being rated are checked. The sum of the scale values of the statements checked, divided by the number of statements checked, gives the score of the child rated.

[1] The manuscript of the original study, describing in detail the process of devising the scales, may be borrowed from the Merrill-Palmer School.

When a sufficient number of children have been rated on the scale, a percentile table can be derived from the results and the score of an individual child can then be interpreted in terms of percentile rank.[1]

MERRILL-PALMER SCHOOL

Progress in Singing Development

Date_____

Name_____ Rated by_____

Directions for Rating:

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure. A few true statements are better than many half true ones.

Scale Value		
9	1. ( )	Recognizes a song upon hearing the opening phrase.
6	2. ( )	Recognizes a song after hearing it played through once.
1	3. ( )	Does not recognize a song (previously taught).
10	4. ( )	Matches a single tone accurately at first attempt.
6	5. ( )	Matches a single tone accurately after several attempts.
2	6. ( )	Cannot match a single tone with any accuracy of pitch.
10	7. ( )	Matches a series of simple tone combinations accurately (at first attempt).
7	8. ( )	Matches a series of simple tone combinations accurately after several attempts.
6	9. ( )	Matches a series of simple tone combinations so that a close resemblance is apparent.
3	10. ( )	Cannot match a series of simple tone combinations.
10	11. ( )	Matches a series of long and difficult tone combinations at first attempt.
7	12. ( )	Matches a series of long and difficult tone combinations so that a close resemblance is apparent.
8	13. ( )	Recognizes when the flow of the melody is upward.
3	14. ( )	Does not recognize when the flow of the melody is downwards.
10	15. ( )	Imitates accurately the calls of things about him (bells, birds, etc.).
3	16. ( )	Makes no attempt to imitate the calls of things about him.
8	17. ( )	Likes to sing songs taught him.
3	18. ( )	Does not like to sing songs taught him.
8	19. ( )	Sings spontaneously as he works or plays.
4	20. ( )	Does not sing as he works or plays.
2	21. ( )	Tries to avoid singing.

[1] Sample copies of the scales are available from the Merrill-Palmer School, 71 East Ferry Avenue, Detroit, Michigan, at five cents each.

8	22. ( )	Sings with a pleasing quality of voice.
4	23. ( )	Sings with a harsh quality of voice.
2	24. ( )	Sings in a monotonous drawl.
2	25. ( )	Mumbles when he is asked to sing.
9	26. ( )	Is critical of his own singing.
5	27. ( )	Is not critical of his own singing.
10	28. ( )	Creates songs with original words and original melody.
9	29. ( )	Can put a phrase or sentence to music.
5	30. ( )	Does not create songs.
4	31. ( )	Does not ask to have music sung or played for him.
8	32. ( )	Asks to have music sung or played that he has heard before.
7	33. ( )	Sometimes expresses the mood of the music in his singing.
8	34. ( )	Recognizes the difference between loud and soft.
5	35. ( )	Sometimes recognizes the difference between loud and soft.
2	36. ( )	Never recognizes the difference between loud and soft.
9	37. ( )	Recognizes the differences between sad and gay.
6	38. ( )	Sometimes recognizes the differences between sad and gay.
3	39. ( )	Never recognizes the difference between sad and gay.
10	40. ( )	Plays little original tunes on an instrument such as water glasses, xylaphone or piano.
5	41. ( )	Not interested in playing tunes on water glass, xylaphone, piano, etc.
4	42. ( )	Indifferent to music period.

MERRILL-PALMER SCHOOL

Development in Rhythmic Ability

Date_____

Name_____ Rated by_____

Directions for Ratings:

Check only those statements which you feel are really true of the child. Do not guess if you are not reasonably sure. A few true statements are better than many half true ones.

Scale Value		
8	1. ( )	Marches to music played in march time.
2	2. ( )	Makes no response to music played in march time.
4	3. ( )	Performs some other rhythmic activity to music played in march time.
9	4. ( )	Recognizes march time when a new march is substituted for one previously taught.
4	5. ( )	Does not recognize march time when a new march is substituted for one previously taught.

6 6. ( ) Can continue rhythm after slight confusion when a new march is substituted for one previously taught.
8 7. ( ) Runs in response to music for running activity.
2 8. ( ) Makes no response to music for running activity.
4 9. ( ) Performs some other rhythmic activity to music for running activity.
6 10. ( ) Recovers rhythm after slight confusion when a new piece of music for running activity is substituted for one previously taught.
3 11. ( ) Cannot recover rhythm lost in changing to a new piece of music for running activity.
10 12. ( ) Immediately recognizes that the music is a running activity when a new piece of music is substituted for one previously taught.
8 13. ( ) Skips to music for skipping activity.
3 14. ( ) Makes no response to music for skipping activity.
10 15. ( ) Immediately recognizes music for skipping when a new piece of music is substituted for one previously taught.
5 16. ( ) Becomes confused when a new piece of music for skipping is substituted for another previously taught.
6 17. ( ) Performs the correct rhythmic activity when he hears the entire piece of music previously taught.
9 18. ( ) Always performs the correct rhythmic activity for music previously taught.
7 19. ( ) Usually performs the correct rhythmic activity for music previously taught.
3 20. ( ) Seldom performs the correct rhythmic activity for music previously taught.
1 21. ( ) Never performs the correct rhythmic activity for music previously taught.
2 22. ( ) Makes no attempt to respond rhythmically to music previously taught.
8 23. ( ) Responds to music which he has not heard before by tapping or clapping.
4 24. ( ) Makes no response rhythmically to music which he has not heard before.
10 25. ( ) Expresses his own ideas of music through creative responses.
4 26. ( ) Has to know the name of the music before he can perform the rhythmic activity.
2 27. ( ) Cannot perform the correct rhythmic activity even when told the name of the music.
5 28. ( ) Performs the correct rhythmic activity as soon as other children lead the way.
5 29. ( ) Becomes confused when there is a change in rhythm.

7	30. ( )	Recovers the rhythmic activity after becoming confused by a change of rhythm in the music.
10	31. ( )	Recognizes similarity of rhythmic patterns in different pieces of music.
9	32. ( )	Can express through rhythmic activity contrasts in intensity of sounds.
4	33. ( )	Cannot express through rhythmic activity contrasts in intensity of sounds.
10	34. ( )	Can express through rhythmic activity different note values.
7	35. ( )	Interprets the mood of the music fairly well.
9	36. ( )	Requests music to express his own rhythmic activities.
10	37. ( )	States definitely the kind of music he wants to express his own activities.
9	38. ( )	Likes to play rhythms on percussion instruments.
3	39. ( )	Has no desire to play rhythms on percussion instruments.
9	40. ( )	Plays rhythmic patterns on instrument so that it coincides with patterns of music played.
7	41. ( )	Sometimes plays rhythmic patterns on instrument so that it coincides with patterns of music played.
10	42. ( )	Creates new rhythmic patterns on percussion instrument.
5	43. ( )	Never creates new rhythmic patterns on percussion instrument.
5	44. ( )	Has greater interest in the mechanical aspects of the instrument than in the music.

AN UNEMPHASIZED FACTOR IN CURRENT THEORIES REGARDING THE TRANSFER OF TRAINING

ERNEST E. BAYLES

Associate Professor of Education, University of Kansas

The title of this paper states fairly the writer's intention. There is no thought of presenting a new theory regarding transfer of training, nor even of presenting new data. Our only purpose is to assist in emphasizing what appears to be, until recently, an unemphasized and almost unrecognized factor in transfer. This factor seems to be sufficiently important for teaching theory that a considerable amount of attention will not be amiss.

Only a brief sketch of past and current interpretations of transfer is necessary here. First is the theory of faculties, in which transfer is assumed to be well-nigh complete. Formal discipline has never recovered from Thorndike and Woodworth's experimentation at the turn of the century. If a faculty were to possess the traits it was supposed to possess, improvement in function must be as large in one field as in another. When this hypothesis was put to the crucial test, nature gave a negative answer. Transfer did not even approach one hundred per cent.

As we see it now, the difficulty with the faculty theory was that it assumed transfer to be automatic; dependent entirely upon conditions posed by the environment. When*ever*, the environment established a condition conducive to transfer, the properly trained individual, according to the theory, was supposed to take advantage of it and act accordingly. If the individual had become good at reasoning or observing, he would reason or observe well, supposedly, every time the situation was such as to call for reasoning or observation. Both experience and experiment combine to prove that such an outcome is never achieved. Education would be far simpler in many ways if such an hypothesis would only work.

In order to replace the faculty theory, Thorndike proposed the theory of "identical elements." Training would transfer when, only when, and in the degree to which, previously learned elements appeared in later situations. This is the reflex arc or S-R bond concept. But here again is a notion of automatic transfer; a notion that the individual will transfer his training when*ever* the environment establishes a favorable set of conditions. After all, such a theory is fundamentally the same as the faculty theory, for all one needs to do is to divide

faculties into sufficiently small units, and behold, one will have identical elements. If transfer does not occur when*ever* the environment sets the stage (that is, whenever a proper stimulus is applied) then the theory breaks down. And it can be, and has been, easily demonstrated that identical elements do not achieve one hundred per cent transfer.

One might argue that the difficulty lies in the fact that we have not yet found the *true* elements. This may well be, but the search for true elements has been followed persistently for many years and it seems to be getting nowhere. Moreover, physicists and chemists have virtually abandoned the hope of finding ultimate elements, and it may be wise for psychologists to do likewise.

Between 1915 and 1920 the argument waxed hot between Thorndike and Judd as to whether transfer was to be obtained through the agency of identical elements or of generalizations. Judd's dart-throwing experiment seemed to show that identical elements in the environment were not enough. It appeared to show the advantages of a "generalized" training; that is, an individual must have been shown during his training that situations seem to work according to discoverable *rules*, and that one must become able to apply the rule (or generalization) before one can take advantage of environmental recurrences and use one's training profitably.

It would seem that the generalization-of-experience theory has some advantage over the identical-elements theory. Long since have Thorndike's pupils reported that he admits the advantage of awareness on the part of the individual of the possibility of transfer. If the individual approaches a new situation with the idea that he may be able to apply ideas, rules, or generalizations previously learned, he will be more likely to interpret the new as being similar to the old, and act accordingly, than if he does not so approach it. This, as I see it, is what training in generalization tends to promote, but does not emphasize sufficiently. The question, whether or not transfer will *always* occur after the individual has learned a given principle or generalization, still seems to be answerable only in the negative.

I believe it just as futile to expect a generalization to spring into action whenever the environment sets the stage as to expect such of a faculty or of an "identical element." The difficulty is that the theory of transfer-by-generalizations still leaves the emphasis upon the environment as the dominant motivating agency; it still interprets transfer as being automatic when the environment sets the stage.

Transfer of a given principle will not be one hundred per cent even when the individual understands the principle thoroughly and has applied it often. For example, an object will float in water if its specific gravity is less than the specific gravity of water. This is a phase of Archimedes' principle, and is well understood by many general science and physics students. But how many will recognize, without being told, that Archimedes' principle applies equally well to objects floating in liquids other than water, and to objects floating in air? How many students will solve correctly problems dealing with flotation in air after having received instruction on flotation in water? Not a large proportion; certainly not one hundred per cent. Some will catch the relationship immediately; others will see it if given a bare hint; others, if the similarity is carefully explained; and possibly others, not at all. Evidently something more than commonness of elements or appropriateness of generalizations is necessary in order to make transfer dependable. In so far as this "something" is missed, the transfer-by-generalizations theory is little better than the identical-elements or even the faculty theory.

If I were limited to a choice among the three, I would choose the generalizations theory, for it demands more specific training than does the theory of faculties, and is at the same time broader and more educative than the theory of identical elements. But since a generalization may be the identical element in question, the principle of transfer is the same in both cases. This similarity has been noted by Sandiford, a pupil of Thorndike, who claims that both theories amount to one and the same thing. Orata, whose analysis of the literature on transfer of training is probably as thorough and as sound as any in existence, makes the following statement:

> The problem . . . is the ability not so much to generalize and apply as to deal with changing ideals and standards, which means the ability and willingness to change the generalizations made previously. . . . This ability to deal with changing principles or generalizations Judd's theory does not provide for, much less does Thorndike's theory of identical elements.[1]

As a result of Orata's summary of all studies of transfer made between 1890 and 1935, he concludes that,

> Forty-seven or nearly thirty per cent show considerable transfer, eighty or nearly fifty per cent show appreciable transfer, fifteen or less than ten per cent show little transfer, six or less than four per cent show no transfer, and

[1] *Ed. Adm. & Superv.*, Vol. XXI, No. 4, 256.

the rest which comprise less than ten per cent show both transfer and interference. And since interference is indicative of transfer of a negative character, it is safe to conclude that all doubts with reference to the possibilities of transfer of training may be cast away.[1]

But while Orata concludes that "the fact of transfer can no longer be doubted," he also concludes (in the same sentence) that "it is not an automatic process." (p. 252)

Now, since all of the current theories which appear to assume that transfer will always occur whenever a particular kind of environment impinges upon a particularly trained individual fail to dispose of the problem satisfactorily, what then? What is the rest of the story? What essential feature of the picture has been omitted?

Bagley seems to hit close to the mark when he introduces the concept of the "conscious ideal." He found little transfer of the "ideal" of neatness from class *A* to other classes, when the ideal was applied to class *A* papers only. But he found considerable transfer after the teacher of class *A* had dwelt upon the desirability of neatness papers prepared for classes other than class *A*. I have heard Judd speak of the necessity of teaching not only the generalization in question, but also of teaching pupils at the same time that opportunities for applying the generalization will probably be met frequently and that they will do well to look for such opportunities. Thorndike and Gates go on record as follows:

> Studies of the transfer of training also have shown that the methods used in guiding the pupils' learning activities have marked effect upon the degree of transfer. The more clearly the crucial element or fact or principle in a situation is brought to the pupil's attention the more readily the same element or fact or principle may be identified in another situation. . . . If a child observes, despite many differences in details in a new mechanical puzzle, that the vital principle is the same as in puzzles previously solved, the solution is more likely to be achieved than when the common principle is not identified.[2]

The following quotation is from Mursell:

> Our general comment on this (the identical-elements) theory is that identical elements between two situations must always be identities which exist *in the learner's mind.* The mere fact that certain English words are derived from Latin will not guarantee transfer, and indeed will not produce it in any way, unless the identity is perceived and responded to by the pupil. And this

[1] *Ibid.*, pp. 242–243.

[2] *Elementary Principles of Education*, p. 104; quoted by Orata, *Ibid.*, p. 245.

absolutely essential recognition, perception, and reaction to identities in various situations, is not a passive affair, which can be counted on to occur.[1]

The point seems to be that individuals must become to a degree conscious or aware of common factors in differing situations—be they elements, generalizations, rules, ideals, meanings, or what. They must *comprehend* the elements *as* common, the generalizations *as* applicable, or the meanings *as* appropriate. They must, somehow or other, *sense* the similarities between later and earlier experiences.

But does this sensing of similarities or relationships need always to be on the level of clear and full consciousness? Apparently not; because many times we act upon the basis of a vaguely felt hunch—intuitively—without being able to state clearly our reasons for so acting. Let me suggest that it seems probable that we transfer previous learning *whenever and wherever* we sense a later experience as being similar, more or less completely, to an earlier one. This "sensing of similarities or relationships" may be vague, or it may reach the level of fully conscious understanding. The more clearly or fully the relationships are sensed or recognized the greater is the likelihood of transfer. And conversely, the less clearly or fully the similarities or relationships are recognized, the less is the likelihood of transfer. It is not enough that the environment set the stage for recognition; the individual is active as well, and the nature of his activity—of his goals and his insights—will be the deciding factor in determining whether or not he will take advantage of the opportunity offered by what confronts him.

Finally, may I point briefly to what this modified view may mean for teaching. If we want our in-school teaching to transfer in the greatest possible degree to out-of-school behavior, it behooves us to make the student acutely aware of the possibilities of transfer. We must bring him, as frequently as possible, to see the relationships between in-school and out-of-school situations. In having such relationships repeatedly brought to his attention, the student will become progressively more sensitive to the existence of common or universal features in behavior, more aware of the possibilities of discovering and using such relationships, more likely to discover and apply them for himself, and thus more apt to transfer previous training into new areas of experience. If we would have training transfer in high degree, we must make the student acutely sensitive to the fact

[1] *The Psychology of Secondary-School Teaching*, pp. 104–105.

that every new experience is likely to have much in common with the old. He must be brought to realize that rules used with success at prior times have likelihood of being successfully used again and again. And he must also be brought to see that new problems may also be solved by old rules slightly modified if only the right ones can be found.

One must be able, of course, to see differences as well as similarities; else needless mistakes will be made. No two sets of experiences are ever exactly alike. Life is forever new. Training which will enable one to meet life adequately—that is, training which will transfer —is merely training in being intelligent . . . thoughtful . . . scientifically-minded.

If we would make training transfer, our teaching must accomplish two objectives more or less simultaneously: First, it must bring the student to understand as many widely useful relationships, principles, or generalizations as possible; second, it must whet the student's realization that his previous training has wide possibilities for transfer, but that transfer is never automatic. It must bring a realization that transfer comes only if and when one senses for one's self that transfer is possible. One becomes able to make the greatest transfer of previous training when, and only when, as a matter of habit, one is on the alert to recognize previous relationships, principles, meanings, or generalizations which may be used in the new situations that are always at hand.

NOTE ON A GRAPHIC METHOD FOR DETERMINING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN GROUP FREQUENCIES

JOSEPH ZUBIN

College of the City of New York

In statistical investigations in education, psychology, economics, and public health, it sometimes becomes necessary to compare the incidence of many factors in two distinct groups. Examples of such comparisons are (1) the validation of items of a test, (2) the comparison of the rates for various diseases in two communities, (3) the comparison of the types of expenditures in two different income groups, (4) the comparison of "passes" and "failures" on a series of tests for "good" and "poor" readers or for "successful" and "failing" salesmen. Ordinarily a considerable amount of computation is involved in these comparisons. The purpose of this paper is to provide a graphic scheme which will obviate the need for any computations. The absolute frequencies will be compared directly and the significance of the difference determined by mere inspection. The method will be illustrated by an application to item analysis.

The usual procedure for selecting the items of a test that differentiate significantly between two discrete groups may be described as follows: The responses of all the individuals to each of the items are tabulated and the incidence of each response in the two discrete groups is compared. For example, if the two groups under comparison are a group of normal individuals and a group of abnormal individuals, their responses will be tabulated as in Table I.

TABLE I.—INCIDENCE OF NORMAL AND ABNORMAL RESPONSES TO ITEM 1

	Yes	No	Total
Normals	a	b	$(a + b)$
Abnormals	c	d	$(c + d)$
Total	$(a + c)$	$(b + d)$	N

In Table I a is the number of normal individuals who responded *Yes* to item 1, b is the number of normal individuals who responded *No*, c is the number of abnormal individuals who responded *Yes*, and d is the number of abnormal individuals who responded *No*. The usual

procedure in determining whether the two groups differ in their responses to item 1 is to compare the proportion of normals who responded *Yes*, $a/(a + b)$, with the proportion of abnormals who responded *Yes*, $c/(c + d)$. The difference between these two proportions is then compared with its standard error to determine its statistical reliability. The critical ratio, C_1, for this comparison can be written:

$$C_1 = \frac{\left[\frac{a}{(a+b)} - \frac{c}{(c+d)}\right]}{\sqrt{\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}}}. \tag{1}$$

This is usually written in percentages or proportions and appears in the familiar form as equation (2).

$$C_1 = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}}} \tag{2}$$

where

$$p_1 = \frac{a}{(a+b)} \qquad q_1 = \frac{b}{(a+b)}$$
$$p_2 = \frac{c}{(c+d)} \qquad q_2 = \frac{d}{(c+d)}$$
$$N_1 = (a+b) \qquad N_2 = (c+d)$$

The hypothesis underlying this comparison is that the normal individuals do not differ from the abnormal individuals in their response to the item in question. From the point of view of the item, the normal and abnormal groups are tentatively regarded as two samples drawn from the same population. If the critical ratio of the difference between the incidence of the *Yes* response in the two samples is low, that is, if it is no larger than the value that could arise by chance between five and ninety-five times in one hundred, the hypothesis that normals do not differ from the abnormals in their *Yes* response to this item is tenable. If, on the other hand, the critical ratio turns out to be large, (so large that it could arise by chance less than five times in one hundred) the tentative hypothesis that the two groups are samples of the same population is rejected. In this case the response is considered to differentiate between the two groups.

Under the hypothesis that the two samples are drawn from the same population, the best estimate of p, the proportion of *Yes* responses in the population, is the weighted average of p_1 and p_2, which will be designated as $p' = [N_1p_1 + N_2p_2]/N$. The standard error of p_1, $\sigma p_1 = \sqrt{p'q'/N_1}$, the standard error of p_2, $\sigma p_2 = \sqrt{p'q'/N_2}$ and the standard error of the difference is

$$\sigma(p_1 - p_2) = \sqrt{\frac{p'q'N}{N_1N_2}}$$

and

$$C_2 = \frac{[p_1 - p_2]}{\sigma(p_1 - p_2)} = (p_1 - p_2)\sqrt{\frac{N_1N_2}{(p'q'N)}}.$$

Resolving this formula to the absolute frequency form, it becomes:

$$C_2 = \frac{(ad - bc)\sqrt{N}}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}. \tag{3}$$

The value of C obtained in this manner is equal to ϕ^2N where ϕ^2 is the square of the Boas-Yulean ϕ which is the four-fold point correlation coefficient.[1] If we square C_2 in equation (3) we obtain

$$C_2^2 = \frac{(ad - bc)^2N}{[(a + b)(c + d)(a + c)(b + d)]} \tag{4}$$

Equation (4) is the expression for Chi square. We shall henceforth replace C_2^2 by Chi square.

The value of C_2 given by equation (4) is not identical with that of C_1 given by equation (2). But it has been demonstrated that C_2 is identical with χ^2, and the value of χ^2 is the best approximation that is now available for the determination of the exact probability of the difference between frequencies. Hence C_2 is to be preferred to C_1, despite the fact that C_1 has a greater vogue in educational and psychological work. The advantage of the newer method results from the fact that $'p'$, the population value of p', can be estimated with greater exactness from the data than either $'p_1$ or $'p_2$, the two population values corresponding to p_1 and p_2, because $'p'$ is estimated from a larger sample. The value of C_2 may be interpreted in terms of the probability table for deviations from the mean in a unit normal distribution,

[1] Yule, G. U.: *An introduction to the theory of statistics.* London: Charles Griffin and Co., 1927, Eighth Edition, p. 217, formula (12).

or in terms of Chi square with one degree of freedom by means of the table provided by Fisher.[1]

COMPUTATION OF C_2 AND CHI SQUARE

When only a few comparisons are to be made the significance of the difference can be obtained most readily by the use of the Edgerton-Paterson Tables,[2] or by the use of the Transformation Function provided by the present writer.[3] Both of these methods require that the obtained frequencies be converted into proportions or percentages before the comparison can be made. When a very large number of comparisons are to be made between the incidence of responses in two groups, the method to be described in this paper offers a considerable saving of time. The first advantage of the new method results from the fact that the frequencies can be compared directly and do not have to be converted into proportions. The second advantage of the new method results from the fact that it eliminates the computation of standard errors and critical ratios.

Usually the exact value of Chi square is not the end sought in the analysis. More often, the only information required is whether the value of Chi square falls above or below the critical value required for statistical reliability, (6.635 for 2×2 fold tables with one degree of freedom).[4] In order to determine whether a given combination of values of a and c in Table I yields a significant value of Chi square, we need to know the critical values of c associated with a given value of a. We wish to know how much c must excel or fall below a in order to produce a statistically significant difference. It can be readily

---

[1] Fisher, R. A.: *Statistical Method for Research Workers.* Edinburgh: Oliver Boyd, 1934, Fifth Edition, p. 89.

The relationship between C_2 and Chi Square is given in Fisher, R. A.: "On the Interpretation of χ^2 from Contingency Tables and the Calculation of P." *Jr. Royal Statistical Soc.*, 1922, p. 91. It is also given in the lecture notes of Harold Hotelling, and has also recently appeared in Irwin, J. O.: "Tests of significance for differences between percentages based on small numbers," *Metron*, Vol. XII, 1935, pp. 83–94.

[2] Edgerton, Harold A. and Paterson, Donald G.: "Table of Standard Errors and Probable Errors of Percentages for Varying Number of Cases." *Jr. Applied Psychology*, Vol. X, No. 3, Sept., 1926, pp. 378–391.

[3] Zubin, J.: "A Transformation Function for Proportions and Percentages." *Jr. Applied Psychology*, Vol. XIX, No. 2, pp. 213–220.

[4] It has become customary in psychological and educational investigations to take $P = .01$, as the critical value of the probability rather than $P = .05$ as suggested by Fisher.

shown that the locus of these critical values of c and a is an ellipse. After simplifying equation (4) and remembering that Chi square can be replaced by $N \cdot \phi^2$, we obtain equation (5)

$$a^2(\phi^2 + r)r + 2ac(\phi^2 - 1)r - c^2(r\phi^2 + 1) - (a+c)rN\phi^2 = 0. \quad (5)$$

Where

$$r = \frac{(a+b)}{(c+d)}.$$

Equation (5) is the equation of an ellipse in variables a and c. The other terms in equation (5) are constants. The value of ϕ^2 is a constant as long as Chi square is kept constant at the minimum level of significance and N remains unchanged. The value of r also remains constant as long as the frequencies in the two criterion groups, $(a + b)$ and $(c + d)$ remain constant. Equation (5) can be reduced to the standard form for the ellipse by two transformations. First, the center of coördinates has to be moved to the point $(a + b)/2$, $(c + d)/2$, in order to remove the first degree terms. Then the axes must be rotated through an angle $\theta = \frac{1}{2}\tan^{-1}\frac{2r(1 - \phi'^2)}{(1 - r^2)}$ in order to remove the cross product term. After these transformations are carried out, equation (5) becomes equation (6).

$$X^2 \cdot \frac{2(n-m)}{r\phi'^2 \cdot N^2} + \frac{Y^2 \cdot 2(n+m)}{r\phi'^2 \cdot N^2 - 1} = 0. \quad (6)$$

Where X^2 and Y^2 are the new variables and $n = 1 + r^2 + 2r\phi'^2$;

$$m = \pm\sqrt{(1 - r^2)^2 + 4r^2(1 - \phi'^2)^2}$$

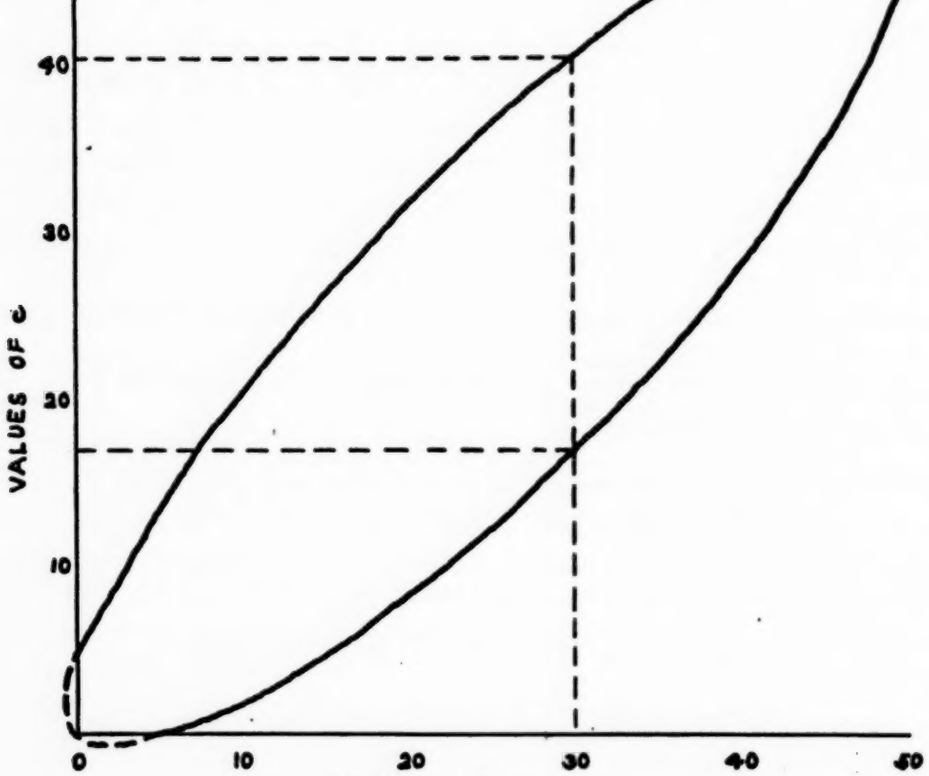
and ϕ'^2 is the critical value of ϕ^2 corresponding to a χ^2 of 6.635.[1]

Equation (6) is the equation of an ellipse whose center is at $(a + b)/2$, $(c + d)/2$ and whose major axis makes an angle of θ with the a axis. The length of the major axis of this ellipse is $k' = \sqrt{2rN^2\phi'^2/(n - m)}$. The length of the minor axis is $k'' = \sqrt{2rN^2\phi'^2/(n + m)}$. After drawing the ellipse according to the above specifications, the area bounded by the circumference of the ellipse contains all the points whose coördinates give rise to values of Chi square that fall below the

[1] Only the positive value of m needs to be considered. The negative value leads to an ellipse that is identical with the ellipse resulting from the negative value of m, except for the interchanging of the ordinates and abcissae.

critical value. The area outside of the ellipse contains all the points whose coördinates give rise to values of Chi square that exceed the critical value. The points on the circumference of the ellipse are the critical points sought. Regarding *a* as the independent variable there will be two values of *c* corresponding to any given value of *a* lying on

GRAPH I
CRITICAL VALUES OF THE ABSOLUTE FREQUENCIES YIELDING SIGNIFICANT DIFFERENCES



the ellipse. These are the two critical values of *c*, the minimal and the maximal, associated with the given value of *a* necessary to yield a value of Chi square above the critical value.

Similar ellipses could be drawn for values of χ^2 other than the critical value of 6.635. All of these ellipses will be concentric and coaxial. They will differ in the lengths of their respective major and minor axes. If all of the ellipses corresponding to the various significant probability levels of χ^2 are drawn, the level of probability of the

difference for any combination of values of a and c can be determined directly from the graph.

An example of the application of this method will be given for the condition, $N = 100$, $(a + b) = 50$. This is the situation when the frequency in the normal and abnormal groups described above are each equal to 50. For this case, $r = 1$, $m = 2(1 - \phi'^2)$, $n = 2(1 + \phi'^2)$, $n - m = 4\phi'^2$, $m + n = 4$. The angle θ through which the ellipse must be turned is 45°. The length of the major axis, k' is 2(35.355) and the length of the minor axis k'' is 2(9.109). The distance between the two foci of the ellipse and the center of the ellipse, k''', is +34.165, since $\frac{1}{2}k''' = \frac{1}{2}\sqrt{k'^2 - k''^2}$. The center of the ellipse is at the point 25, 25. We now have all the data required for drawing the ellipse. Table II has been prepared from the ellipse by reading off all the values of a and c that lie on the ellipse. Graph I shows the ellipse and the two critical values of c that correspond to the value of $a = 30$. The dotted vertical line at $a = 30$ cuts the ellipse at points $c = 17$ and 41. Hence when $a = 30$, c must be 17 or less, or 41 or more, in order to yield a significant difference.

A simple method for drawing the ellipse is to locate its center on a sheet of coördinate paper at the point P whose coördinates are $(a + b)/2$, $(c + d)/2$. At this point draw a line at an angle θ to the horizontal axis, a. This line is the line of the major axis of the ellipse. Erect a perpendicular to the line of the major axis at the center of the ellipse. On this perpendicular lay off the distances $\pm k''/2$, the length of the semi-minor axis. On the line of the major axis lay off the distances of $\pm k'/2$, the length of the semi-major axis. Also lay off the major axis the distances $\pm k'''/2$, the distances of the two foci from the center. Place thumb-tacks at the top of the minor axis, and at the two foci. Draw a string around these three points as verteces. Wind a pencil or a stylus through a loop in this string and twist it until the proper tautness is attained. Remove the thumb tack from the point at the top of the minor axis and run a pencil tautly along the string. The pencil will describe the desired ellipse. Make sure that this ellipse passes through the ends of the major and minor axes. If it does not, the length of the string must be changed until the curve passes through these points.

A still further consideration of equation (4) yields the condition under which $(ar - c)$ always yields a value of Chi square that exceeds the critical criterion. Equation (4) can be rewritten as follows:

$$(ar - c) = \sqrt{Nr\chi^2 p'' q''} \qquad (4')$$

where

$$p'' = \frac{(a+c)}{N}, \qquad q'' = \frac{(b+d)}{N}, \qquad p'' + (q'' = 1$$

Now the maximum value that $p''q''$ can attain is $\frac{1}{2}$.

TABLE II.—CRITICAL VALUES FOR THE COMPARISON OF FREQUENCIES IN TWO GROUPS OF FIFTY INDIVIDUALS

$X - a$	Higher frequency, c'	Lower frequency, c''	a	Higher frequency, c	Lower frequency, c''
0	6	..	26	39	13
1	9	..	27	40	14
2	10	..	28	40	15
3	13	..	29	41	16
4	15	..	30	42	17
5	16	..	31	43	18
6	17	..	32	43	19
7	18	..	33	44	20
8	19	1	34	45	21
9	21	1	35	46	22
10	23	1	36	46	23
11	24	2	37	47	24
12	25	2	38	47	25
13	26	3	39	48	26
14	27	4	40	48	27
15	28	4	41	49	29
16	29	5	42	49	30
17	30	6	43	50	31
18	31	7	44	50	33
19	32	7	45	..	34
20	33	8	46	..	35
21	34	9	47	..	37
22	35	10	48	..	39
23	36	10	49	..	40
24	37	11	50	..	43
25	38	12			

Hence $(ar - c) \not> \frac{1}{2}\sqrt{Nr\chi^2}$. We can now set $\chi^2 = \chi'^2$ the critical value of Chi square and determine the value of $(ar - c)$ necessary to produce a Chi square of 6.635 when $p''q'' = \frac{1}{2}$. Whenever $(ar - c)$ exceeds this value, χ^2 exceeds the critical value, and hence the difference is significant. The line on which the points that satisfy the condition $(ar - c) = \frac{1}{2}\sqrt{Nr\chi'^2}$ lie, has the slope $1/r$ and an intercept of $\pm\frac{1}{2}\sqrt{N\chi'^2/r}$ and can be readily drawn.

For the example described above, ($N = 100$, $r = 1$), $(a - c) = \frac{1}{2}\sqrt{100/6.635} = 12.8$. Hence whenever $(a - c)$ exceeds 13, the difference must be significant. The area in which the points whose coördinates satisfy this condition lie can be determined by drawing a line at the point 12.8, 0 with a slope parallel to the major axis, since $r = 1$. The area to the right of this line contains the points for which the differences are always significant.

If there are more than two choices to each item, the formula is still applicable. Table III shows the results for a three choice item.

TABLE III

	Yes	No	?	Total
Normal	a	b	c	$(a + b + c)$
Abnormal	d	e	f	$(d + e + f)$
Total	$(a + d)$	$(b + e)$	$(c + f)$	N

In this table we can determine whether a is significantly different from d in the following equation.

$$X^2 = \frac{(a - d)^2 N}{[(a + b)(b + c + e + f)(a + b + c)(d + e + f)]} \qquad (7)$$

This value of Chi square is equivalent to the value that would be obtained when each response is pitted against the other two responses combined.[1]

The graphic determination of the significance of the difference between frequencies that has been described above has been limited to one specific value of N. We can generalize the solution for all values of N by dealing with proportional instead of absolute frequencies. This will necessitate the computations of the proportions corresponding to each absolute frequency in the various 2 × 2-fold tables to be examined. But once these proportions are obtained,

TABLE IV

	Yes	No	Total
Normals	p_1	p_2	$p_1 + p_2$
Abnormals	p_3	p_4	$p_3 + p_4$
	$p_1 + p_3$	$p_2 + p_4$	1.000

[1] Fisher: *Op. cit.*, p. 105.

the significance of their difference can be determined from the graph. Converting Table I into proportional frequencies we obtain Table IV. where

$$p_1 = \frac{a}{N}, \qquad p_2 = \frac{b}{N}, \qquad p_3 = \frac{c}{N}$$

and

$$p_4 = \frac{d}{N}$$

Equation (6) can be rewritten in terms of the proportional frequencies into equation (8) by letting $X^2/N^2 = \alpha^2$ and $Y^2/N^2 = \beta^2$

$$\frac{2(n-m)}{r\phi^2}\alpha^2 + \frac{2(n+m)}{r\phi^2}\beta^2 - 1 = 0 \tag{8}$$

and

$$r = \frac{N_1/N}{N_2/N} = \frac{p_1 + p_2}{p_3 + p_4},$$

n, m, and ϕ^2 have the same values as before.

The semi-major axis of the ellipse, $k' = \sqrt{\frac{r\phi^2}{2(n-m)}}$. The semi-minor axis $k'' = \sqrt{\frac{r\phi^2}{2(n+m)}}$ and the center of the ellipse is at the point $\frac{p_1 + p_2}{2}, \frac{p_3 + p_4}{2}$. In terms of r, the center is located at $\frac{r}{2(1+r)}, \frac{1}{2(1+r)}$. It should be noted that equation (8) has only two parameters, r and ϕ^2, since n and m are functions of r and ϕ^2 and of no other parameters. We can therefore keep r constant and draw ellipses for various values of ϕ^2. These ellipses will correspond to various degrees of significance. Thus, the ellipse for $\phi^2 = .10$ will correspond to a χ^2 of 2.5 when $N = 25$, or a χ^2 of 5 when $N = 50$, and so on. All of these ellipses belong to the same family, that is, their axes differ only in size but not in direction. The drawing of these ellipses can be done mechanically by increasing the size of the axes step by step in accordance with the increase in the value of ϕ^2. A new system of such ellipses is required for each value of r. Although r can

vary from 0 to 1, it is only necessary to present different values of the ellipse for r from .00 to .50, since the ellipses are symmetrical about $r = .50$. If r is taken in steps of .01, it would be necessary to prepare fifty such tables. These tables will be exactly accurate when $N \leq 100$, and will probably be approximately accurate for larger values.

A somewhat easier method of providing a generalized solution for proportional frequencies is afforded by the t transformation method.[1] In this method, however, we must deal with percentages based upon the frequencies in the two criterion groups. Letting $P_1 = a/N_1$ and $P_2 = c/N_2$, where N_1 and N_2 are the frequencies in the two criterion groups, we can transform these values to t values and obtain

$$\frac{t_1 - t_2}{\sqrt{N/N_1N_2}} = C, \text{ the critical ratio for the difference.} \qquad (9)$$

Hence

$$(t_1 - t_2) = C\sqrt{\frac{N}{N_1N_2}}.$$

Now

$$(t_1 - t_2) = 2\{\sin^{-1}\sqrt{P_1} - \sin^{-1}\sqrt{P_2}\}$$

or

$$\sin\frac{C}{2}\sqrt{\frac{N}{N_1N_2}} = \sqrt{P_1 - P_1P_2} - \sqrt{P_2 - P_1P_2}.$$

Let

$$\sin\frac{C}{2}\sqrt{\frac{N}{N_1N_2}} = K.$$

By simple algebraic transformation we obtain

$$P_1^2 + P_2^2 + (4K^2 - 2)P_1P_2 - 2K^2(P_1 + P_2) + K^4 = 0. \qquad (10)$$

Equation (10) can be transformed to the typical equation for an ellipse by shifting the center of coördinates to $P_1 = .50$, and $P_2 = .50$ and by rotating the axes through an angle of 45°. After performing these transformations we obtain equation (11)

$$\frac{X^2}{\frac{2 - K^2}{8}} + \frac{Y^2}{\frac{(2 - K^2)}{8} \cdot \frac{K^2}{(1 - K^2)}} - 1 = 0 \qquad (11)$$

[1] Zubin: *Op. cit.*

The length of the semi-major axis is $\sqrt{\frac{2-K^2}{8}}$, the length of the semi-minor axis is $\sqrt{\frac{(2-K^2)K^2}{8(1-K^2)}}$ and the distance between the center and the two foci is $\pm\sqrt{\frac{(2-K^2)(1-2K^2)}{8(1-K^2)}}$.

The ellipse can now be drawn. The value of K is dependent only upon the size of the frequency in the two criterion groups. In order to make the results of this graph comparable with those of Graph I, we take $C = \sqrt{6.635} = 2.58$ and

$$K = \sin \frac{2.58}{2}\sqrt{\frac{N}{N_1N_2}} = \sin 1.29\sqrt{\frac{N}{N_1N_2}},$$

when $N = 100$, $N_1 = 50$, $N_2 = 50$, $K = \sin 0.258 = 0.255$.

It should be noted, however, that P_1 and P_2 are expressed in terms of N_1 and N_2, and not in terms of N. Hence, the values of P_1 and P_2 must be computed anew for each new value of N_1 and N_2. But it seems preferable to follow this method instead of computing the fifty charts required by the previous method, equation (8). In the present method one chart will be sufficient to cover all the possibilities. But this chart must contain ellipses for the different possible values of K. All of these ellipses will be concentric and coaxial and can be readily drawn.

Another example of the application of the graphic method for the solving of a problem of the significance of the difference between frequencies is afforded by the comparison of matched groups. When the matching is done by pairs, the results can be analyzed in the following manner. Let us suppose that a group of normal individuals have been matched with a group of abnormals on several factors. The responses of these two groups to a test can be analyzed in accordance with Table IV.

This table is read as follows: There were forty-five normal individuals who gave the *Yes* response to the item under analysis, and their correspondents the abnormal group also gave the *Yes* response. There were forty-three normal individuals giving the *Yes* response whose correspondents gave the *No* response, and so on.

If the normal individuals had given in every case the same response that their abnormal correspondents had given, only the long diagonal of the table would contain any frequencies. The relationship between

the normal and the abnormal groups would be a maximum, and each cell would have a value of Chi square that would be far in excess of chance. Hence in the selection of items that differentiate between the normal and abnormal groups, such an item would be discarded. Only the items that give rise to long diagonal cells with values of Chi square that fall short of the critical value, would be retained. In terms of the graphic solution we are utilizing, we would attempt to find the points that lie within the area of the ellipse. In the case of the cells that are not in the long diagonal, only those for which the value of χ^2 is large indicate differential responses. However, in this problem, the marginal frequencies vary from cell to cell, only those cells that lie in the same row having the same marginal row frequency. That is, the value of r will vary from row to row. In order to be able to determine at a glance whether the frequency in a given cell is significant or not, it is necessary to prepare graphs of the equation of the ellipse for different values of r as described previously, or utilize the adaptation of the t method.

TABLE IV

		Normals			Total
		Yes	No	?	
Abnormals	Yes	45	30	33	108
	No	43	29	28	100
	?	34	16	21	71
	Total	122	75	82	279

Another important use for the graphic solution of the significance of the difference between two frequencies is found in connection with the computation of the tetrachoric correlation coefficient. The computation of these coefficients usually consumes a great amount of time, even when the graphs provided by Thurstone[1] and his co-workers are utilized. Very often the correlation coefficient is found to be statistically unreliable after the laborious process has been carried out. To safeguard against such occurrence, it is advisable to compute

[1] Thurstone, L. L., *et alia: Computing Diagrams for Tetrachoric Correlation Coefficients.* University of Chicago Bookstore, April, 1933, pp. 57.

the value of Chi square for the table before the computation of the tetrachoric correlation coefficient is begun. If the value of Chi square indicates a relationship that is statistically reliable, then no matter how low the correlation may be, it is still significant. By utilizing the graphic method suggested in this paper, the value of Chi square for a given table can be determined with very little effort. The computation of unreliable coefficients can thus be avoided.

Furthermore, for some investigations, the Boas-Yulean correlation coefficient may replace the tetrachoric correlation coefficient. The value of the square of the Boas-Yulean ϕ can be read off from the graphs discussed in this paper. After extracting the square root of the value obtained from the graph, the sign of the coefficient must be determined by inspecting the table.

SUMMARY

It has been pointed out that the formula for the critical ratio between frequencies that is most often used in psychological and educational studies can be replaced by another which has certain statistical advantages. The square of the value of the critical ratio obtained by the second method is identical with the value of Chi square and is equal to $N \cdot \phi^2$ where ϕ is the Boas-Yulean correlation coefficient. The equation for Chi square in terms of the two frequencies under comparison is an ellipse that can be readily drawn. The area enclosed within the ellipse contains the points whose coördinates give rise to the values of Chi square that fall below the critical value needed for significance. The area outside of the ellipse contains the points whose coördinates give rise to the values of Chi square that exceed the critical value.

This method is applicable to all multiple choice items. When the number of choices exceed two, the value of Chi square obtained for each choice is equivalent to the value of Chi square that would be obtained for each choice when it is pitted against all the other choices combined.

INFLUENCE OF ORDER OF PRESENTATION OF ITEMS IN RANKING

PERCIVAL M. SYMONDS

Teachers College, Columbia University

It has long been recognized that the order of presentation of material from which a choice or selection is to be made has been a factor in psychological experimentation. The careful investigator in a choice or selection experiment will repeat his work altering the order of presentation so as to eliminate the influence of this factor. To the writer's knowledge only two studies of this factor have been previously made, both by C. O. Mathews.[1] Mathews found for fourth-, fifth-, and sixth-grade pupils that a given answer on a two-response test is marked 3.2 per cent more often when it is printed to the left of its alternative than when printed to the right, and 33.8 per cent more often when it is printed above rather than below. In a second study[2] Mathews discovered that a similar effect was true in a five-response interest questionnaire.

An opportunity to study the strength of the influence of order of presentation on choice arose in a study of the problems and interests of adolescents in which the data were secured in the form of ranking of fifteen printed items. Rankings were secured from two hundred twelve boys and two hundred forty-seven girls in the junior and senior high schools of Tulsa, Oklahoma,[3] and rankings of the same fifteen items presented in reverse order were obtained from one hundred sixty-one boys and one hundred ninety-two girls in the same school system. Table I shows the distributions of these pupils by grade.

The fifteen items to be ranked were presented on a single sheet in ordinary paragraph form so that each item started directly under its predecessor. Directions for ranking asked that the first three in order be selected first by writing their numbers in printed boxes on the ranking sheet, that the last three in order be selected next, and then

[1] Mathews, C. O.: "The Effect of Position of Printed Response Words upon Children's Answers to Questions in Two-Response Types of Tests." *Journal of Educational Psychology*, Vol. XVIII, 1927, pp. 445–457.

[2] Mathews, C. O.: "The Effect of the Order of Printed Response Words on an Interest Questionnaire." *Journal of Educational Psychology*, Vol. XX, 1929, pp. 128–134.

[3] Through the courtesy of Dr. J. Thomas Wade, Cleveland Junior High School, Tulsa, Oklahoma.

that the remaining items be ranked in groups of three by writing their numbers in their respective boxes.

TABLE I.—DISTRIBUTION OF PUPILS BY GRADE

Grade	When items were presented in original order		When items were presented in reverse order	
	Boys	Girls	Boys	Girls
7	15	19	18	15
8	32	33	28	39
9	64	69	66	66
10	32	53	17	27
11	37	34	11	24
12	30	35	15	15
Post graduate	1	2		
Not stated	1	2	6	6
Total	212	247	161	192

Average ranks for each item and the differences in rank from the items in regular and reversed order were computed.

TABLE II.—AVERAGE RANK OF FIFTEEN ITEMS WHEN PRESENTED IN ORIGINAL AND REVERSE ORDER, AND DIFFERENCE BETWEEN THESE AVERAGE RANKS

Item	Average rank		Difference
	Items presented in regular order	Items presented in reverse order	
1	6.205	7.548	1.343
2	9.686	10.551	.865
3	8.236	9.086	.850
4	6.221	8.047	1.826
5	9.126	9.317	.191
6	7.654	7.890	.236
7	6.958	7.539	.581
8	6.875	7.311	.436
9	8.190	8.327	.137
10	7.221	6.440	− .781
11	6.964	6.510	− .454
12	10.232	9.942	− .290
13	9.334	8.192	−1.142
14	8.682	6.738	−1.944
15	8.364	6.581	−1.783

The linear and quadratic curves of best fit were computed for those points by the method of least squares with the following results:

$$y = -0.220x + 1.764$$
$$y = -0.0106x^2 - 0.048x + 1.266$$

Inspection of the results leads one to the conclusion that the relation between tendency to overrank and order of ranking is linear. The coefficient of the x^2 term in the quadratic is very small and the curvation of the quadratic is slight. The linear equation shows $y = .000$ when $x = 8$ indicating that there is no deviation for the middle rank. It would seem that the strength of the tendency to overrank items coming early on the list and to underrank items coming late on the list is directly proportional to the distance of the item from the middle item.

By a translation of axes the above equation may be written $y = -0.220x$ where $x = 0$ is the middle item in any series. From this simple equation the distortion in rank for any number of items may be computed. For instance, if there are two items to be ranked, for the first item $x = -\frac{1}{2}$ and $y = .11$. That is, if the average rank of the first item is r then the average rank when the order of the two items is reversed is $r - .11$ and the most probable value of the average rank of the first item is $r - \frac{1}{2}$ of .11 or $r - .055$.[1] Similar computations may be made for the ranking of any number of items.

There is at least one factor which destroys the generality of the above relationship. The distortion in ranking due to order of presentation of the items is a function of the method of presentation. Mathews, for instance, in the paper referred to above found the influence of order much greater when items were presented vertically than when presented horizontally. Order of presentation should have practically no effect when the ranking is to be accomplished by sorting cards.

Maturity on the other hand which might be thought to influence the distortion in ranking which is due to the order of presentation of the items apparently has little influence. The rankings of two hundred sixty-six students at the Emporia, Kansas State Teachers College when the material was presented in the original order was compared

[1] I have computed from Mathews' data that the difference in average rank for his two item situation was .032 in the case where material was horizontally presented and .338 where vertically presented against a difference of .110 computed from my data.

with the rankings of two hundred fifty students at the same institution the next year when the material was presented in the reverse order. The results were almost identical with those obtained from the high school group, the relation again being a linear equation $y = -.209x + 1.693$. Evidently the influence of order of presentation of items is not markedly affected by age or maturity.

The influence of order of presentation on the standard deviation of the rankings for each item was investigated. No consistent trend was evident. These data lead to the conclusion that the variability of ranking is not a function of the order of presentation of the material. This may be interpreted as saying that there was no tendency for greater discrimination to be shown in ranking the earlier items of a list.

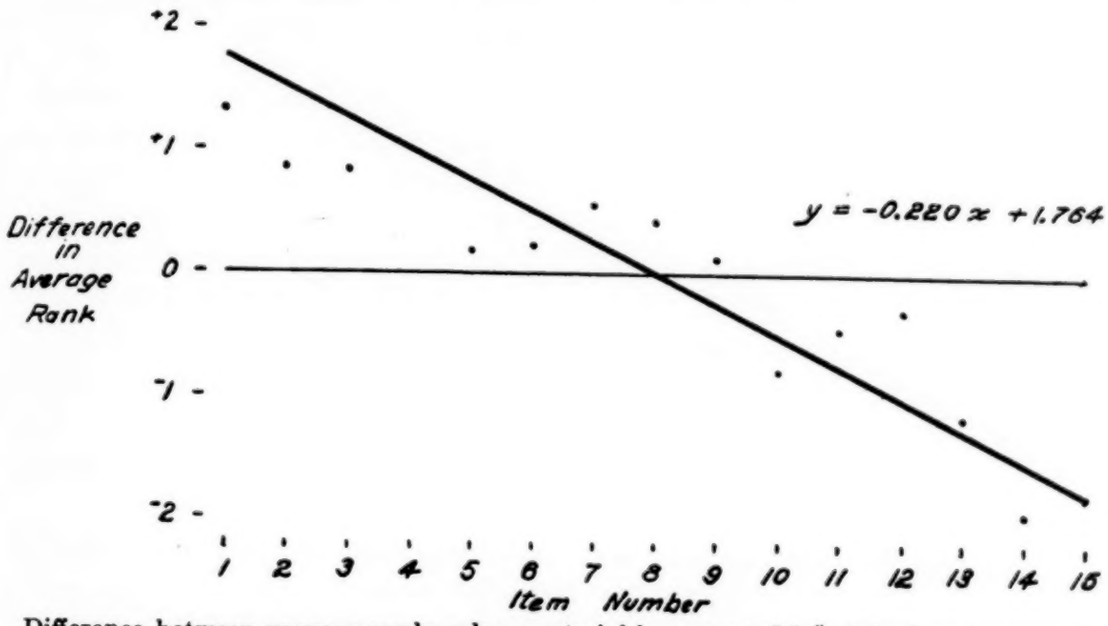
The following tables may be used in correcting for order of presentation in ranking. These tables may be used in correcting rankings obtained from high-school pupils and may be taken as a first approxi-

TABLE III.—CORRECTION VALUES TO BE APPLIED TO AVERAGE RANKING WHEN THERE ARE AN ODD AND AN EVEN NUMBER OF ITEMS TO BE RANKED

When an odd number of items are to be ranked		When an even number of items are to be ranked	
7th item preceding.		7th item preceding.	
middle item	+.770	middle item	+.715
6th	+.660	6th	+.605
5th	+.550	5th	+.495
4th	+.440	4th	+.385
3rd	+.330	3rd	+.275
2nd	+.220	2nd	+.165
1st	+.110	1st	+.055
middle item	.000		
1st item following.		1st item following.	
middle item	−.110	middle item	−.055
2nd	−.220	2nd	−.165
3rd	−.330	3rd	−.275
4th	−.440	4th	−.385
5th	−.550	5th	−.495
6th	−.660	6th	−.605
7th	−.770	7th	−.715

mation until further studies make refinements to care for method of presentation. These tables are one-half the linear regression values and indicate the amount to be added or subtracted to give the most probable average rank.

For instance, if there are five items to be ranked, +.220 should be added to the rank of the first item. If four items are to be ranked, +.165 should be added to the rank of the first item.



Difference between average ranks when material is presented in original and reversed order.

RELATION OF HANDEDNESS, EYE-SIGHTING AND ACUITY DOMINANCE TO READING[1]

ARTHUR I. GATES AND GUY L. BOND

The article reports the results of the relations of handedness, eye dominance for sighting, visual acuity and combinations of these to reading in the case of three, and in some cases four, groups of children, as follows: (1) A group of sixty-five retarded readers with a mean age of 8.61 years and (2) a control group of sixty-five normal readers equivalent in age, intelligence, number of years in school and socio-economic background,[2] (3) a group of fifty-five first-grade pupils measured repeatedly during the year, and (4) a group of fifty-seven first-grade pupils studied during the first half of the year. The first two groups reveal the relationships after several years of school experience, the second pair during the initial stages of learning.

Handedness was determined by five tests as follows: Using a hammer, spinning a top, cutting with scissors, picking up small articles, and throwing a ball. Eye dominance was determined by using the Parson's Manoptoscope and by one or more tests of sighting with a pencil or finger. Visual acuity was measured by the Snellen "E" chart, for distant vision, and the Jaeger chart for near. The acuity test slides of the Betts Telebinocular were also used.[3]

Table I shows the average attainments on different tests secured by pupils showing different degrees of handedness, or more exactly by those testing "right" in all five, or in one to four tests out of five, or left-handedness in all five. It will be seen that this table shows no consistent trend. Among the normal readers, the pupils clearly right-handed have somewhat better scores, whereas among the poor readers the reverse is true. The mixed-hand dominance groups are between in both groups. The number of the left and mixed-handed cases, is, of course, too small to give reliable data. In this table and in others following, measures of statistical reliability have not been

[1] The studies reported in this article were made possible by support received from the Works Progress Administration through the agency of the Board of Education of New York City.

[2] Further details concerning these two groups will be found in Bond, G. L.: *The Auditory and Speech Characteristics of Poor Readers.* Teachers College Contributions of Education, No. 657, 1935.

[3] For details concerning the tests of visual functions see Fendrick, P.: *Visual Characteristics of Poor Readers.* Teachers College Contributions to Education, No. 656, 1935.

computed, since the consistency from group to group is even more significant and can be discerned by casual inspection.

TABLE I.—MEAN SCORES OF PUPILS OF DIFFERENT DEGREES OF HAND DOMINANCE

Handedness	Number of pupils	Reading grade	Word pronunciation	Mispronunciation	Per cent reversals	Per cent wrong beginnings
Normal Readers						
Five right	58	3.9	64	9.2	7.0	15
One to four right	4	3.4	60	11.1	6.5	17
Five left	2	3.5	58	10.9	8.1	14
Poor Readers						
Five right	57	2.0	23	56	9.2	17.6
One to four right	5	2.1	21	59	10.3	13.1
Five left	2	2.2	44	25	8.5	8.5

Among the first-grade pupils, fifty out of fifty-five were graded as right-handed and five as left, in tests given about two months after the opening of school to one group. In two tests of reading given at mid-year and in three at the end of the year, the two groups were almost identical, and in tests of word pronunciation the percentage of reversals errors were ambiguous, as shown in Table II.

TABLE II.—SCORES IN READING ABILITY AND PERCENTAGES OF REVERSALS BY FIRST-GRADE PUPILS OF DIFFERENT HAND DOMINANCE

Handedness	Reading vocabulary mid-year	Paragraph reading mid-year	Word recognition end year	Sentence reading end year	Paragraph reading end year	Reversal errors test 1, per cent	Reversal errors test 2, per cent
Right	13	22	2.3	2.4	2.4	27	17
Left	11	19	2.2	2.4	2.5	37	8

The scores in various tests of pupils showing right, left or uncertain eye dominance for the older children are shown in Table III and for

one of the first-grade groups in Table IV. There is no consistent relationship among these factors in these three groups.

TABLE III.—MEAN SCORES FOR PUPILS OF RIGHT, LEFT OR MIXED EYE DOMINANCE

Eye dominance	Number of pupils	Reading grade	Word pronunciation	Total mispronunciations	Per cent reversals	Per cent wrong beginnings	IQ
Normal Readers							
Right	32	3.9	63	11	5.9	10.6	101.2
Left	28	3.8	63	10	7.2	13.5	101.8
Mixed	4	4.2	66	0	0	16.0	110.7
Poor Readers							
Right	32	2.1	25	50	7.5	17	99.1
Left	30	2.1	23	55	10.0	16	101.1
Mixed	2	1.5	12	141	15.0	12	97.5

TABLE IV.—MEAN SCORES FOR FIRST-GRADE PUPILS OF RIGHT, LEFT OR MIXED EYE DOMINANCE

Eye dominance	Number of pupils	Reading vocabulary mid-year	Paragraph reading mid-year	Average of three reading tests. End year	Reversal test 1	Reversal test 2
Right	29	12.8	22	2.37	35	17
Left	18	11.2	21	2.35	25	15
Mixed	8	12.9	21	2.33	26	21

Table V shows the scores in various tests of the older pupils when grouped according to the eye showing superior visual acuity in near vision by the chart. Differences in both groups may be seen by inspection to be unimportant and unreliable.

Table VI shows the number of errors made in a special pronunciation test including many reversible words by the older pupils having the same and different lateral dominance of hand and eye. Among the normal readers, the cases of mixed dominance show larger per-

TABLE V.—MEAN SCORES OF PUPILS OF SUPERIOR RIGHT OF LEFT OR IDENTICAL EYE ACUITY

Superior eye	Number of pupils	Reading grade	Word pronunciation	Per cent reversals	Per cent wrong beginnings	IQ
Poor Readers						
Right	8	2.01	23	15	8	99.6
Left	10	2.06	29	11	28	102.2
Same	46	2.06	23	9	16	100.4
Normal Readers						
Right	7	3.65	59	3.7	9	100.4
Left	9	3.75	60	7.3	18	100.4
Same	48	3.85	64	5.0	13	102.5

centages of reversals than the right-eyed-right-handed pupils but the trend among the poor readers is the reverse.

TABLE VI.—SCORES ON PRONUNCIATION TESTS OF PUPILS SHOWING DIFFERENT COMBINATIONS OF EYE AND HAND DOMINANCE

Combination	Number of pupils	Total errors	Per cent reversals
Normal Readers			
Right eye—right hand	27	11.2	3.2
Left eye—right hand	23	7.5	5.0
Other mixed combinations	14	8.6	6.5
Poor Readers			
Right eye—right hand	27	46.7	7.0
Left eye—right hand	28	51.4	6.5
Other mixed combinations	9	49.3	6.0

Table VII shows the scores made on two *different* tests of reversal tendencies of which the first is Test VIII—2 and the second Test VIII—3, of the Gates Reading Diagnostic Tests (1935 edition). Each of these tests was given at the beginning, middle, and end of the year to group I (the group previously mentioned) and at the beginning

and middle of the year to forty-six pupils from the other first-grade group. The results for these groups combined comprise Table VII.

TABLE VII.—PERCENTAGE OF REVERSAL ERRORS MADE BY FIRST-GRADE PUPILS GROUPED ACCORDING TO HAND AND EYE COMBINATIONS

Combination	Number of pupils	Recognition of words seen		Recognition of words heard	
		Test 1	Test 2	Test 1	Test 2
Right hand—right eye	39	5.1	6.0	4.5	4.5
Right hand—left eye	24	4.8	4.8	4.5	3.6
Left hand—left eye	2	3.9	5.7	3.9	5.3
Left hand—right eye	3	6.2	4.8	5.7	5.7

TABLE VIII.—MEAN SCORES IN SEVERAL TESTS OF PUPILS OF VARIOUS COMBINATIONS OF EYE DOMINANCE (AIMING TESTS) AND EYE ACUITY

Combination	Number of pupils	Silent reading grade	Perception, G1	Perception, G3	Per cent reversals
Normal Readers					
Right aim—right acuity	4	3.8	11.5	13.0	2.3
Right aim—same acuity	23	4.0	19.7	11.4	3.8
Left aim—left acuity	6	4.0	21.2	12.2	22.4
Left aim—same acuity	18	3.8	18.6	13.1	2.8
All others, mixed	13	3.9	19.3	12.7	3.0
Poor Readers					
Right aim—right acuity	5	2.1	14.6	9.0	8.8
Right aim—same acuity	21	2.6	14.6	9.2	7.4
Left aim—left acuity	5	1.9	13.6	10.4	16.2
Left aim—same acuity	21	2.1	17.0	9.5	10.8
All others, mixed	12	2.0	14.5	9.7	10.4
Above Data for Two Groups Combined and Certain Categories Consolidated					
Right aim—right or same acuity	52	3.3	17.8	10.3	5.5
Left aim—left or same acuity	50	2.9	17.6	11.3	9.1
Mixed	25	3.0	16.8	11.2	6.6

It will be noted that no group shows the reversal tendency reliably greater than the right-handed-right-eyed group.

The older pupils may be divided into several groups according to the relationship of eye-dominance, on the bases of aiming tests, and eye-superiority in acuity. Table VIII gives some of the data when acuity is measured by the near-vision chart. The "mixed groups" include all possible combinations except those listed separately in the table. Tests G1 refers to the Gates Perception of Geometrical figures; Test G3, to Gates Perception of Digits; and Percentages of Reversals to the pronunciation of reversal words test in the Gates diagnostic series.

For the first-grade children similarly grouped are given the scores in a test of perception of digits, and three tests of reversal tendencies in addition to the reading grade based on the mean grade scores obtained from the three Gates Primary Reading Tests given at the end of the year.

The scores for the several groups of the older children are similar except for the relatively high percentage of reversals made by those having left-eye aiming dominance and left-eye superiority in acuity. This tendency does not appear, however, in the results for the first-grade pupils.

TABLE IX.—SCORE IN SILENT READING, PERCEPTION, AND VARIOUS TESTS OF REVERSAL TENDENCIES FOR FIRST-GRADE PUPILS GROUPED ACCORDING TO EYE (AIMING) DOMINANCE AND ACUITY

Combination	Number of pupils	Reading grade	Perception	Per cent reversal, test 1	Per cent wrong order test 1	Per cent reversal, test 2	Per cent wrong order test 2	Per cent reversal, test 3	Per cent reversal, test 4
Right aim—right acuity	6	2.2	26	3	4	5	11	6	23
Right aim—same acuity	17	2.5	37	8	12	9	16	16	25
Left aim—left acuity	3	2.6	41	3	3	5	5	29	17
Left aim—same acuity	11	2.3	32	12	15	9	15	12	19
All others, mixed	15	2.4	33	9	12	6	12	28	41
Above Data Consolidated									
Right aim—right and same acuity	23	2.4	34	7	10	8	15	14	24
Left aim—right and left acuity	14	2.4	34	10	13	8	12	16	19
Mixed	15	2.4	33	9	12	6	12	28	41

In summary it may be said that the data obtained from groups of first grade pupils, older normal readers, and older reading problem

cases, show no consistent tendency for eye dominance, single eye superiority in acuity, hand dominance, or any combination of these to be related to achievement in reading, word pronunciation, reversal errors, or visual perception of various items. A survey of the tables will show a remarkable similarity in the number of pupils in the various categories in the normal and defective reading groups. This in itself is evidence that eye and hand dominance have little to do with reading difficulties. The data, in general, are in harmony with those published by Woody and Phillips, Gates, Kirk, Teegarden, Witty and Kopel[1] and others.

[1] A very complete bibliography of studies on this topic will be found in Witty, P. A. and Kopel, D.: "Sinistral and Mixed Manual-Ocular Behavior in Reading Disability." *Journal of Educational Psychology*, February, 1936, pp. 132–134.

A FORMULA FOR INTERCORRELATIONS AMONG MULTI-SCORES

ELMER B. ROYER

Oklahoma Agricultural and Mechanical College

I. INTRODUCTION

Of late, it has become good practice to score personality and interest blanks a number of times, with a different key each time, and thus to obtain a number of scores from the same pattern of responses. After the blanks have been scored in this manner, the intercorrelations are computed and reported in the literature, presumably as evidence of the relationship existing between the traits which the scores are supposed to represent.

As Lorge[1] has pointed out, the correlations may be due partially to the relationships existing among the keys. In order to show this for the Bernreuter Personality Inventory, Lorge computed the correlations among the keys. Since each key gives separate weights for the "yes," "no," and "?" responses, he computed three sets of intercorrelations. He found the intercorrelations of the "yes" weights and those of the "no" weights to be similar to the intercorrelations of the total trait scores. The correlations of the "?" weights are much smaller in magnitude, though still in the same direction.

Since there is only one set of intercorrelations of the total trait scores, logically, any method which purports to show how much of this correlation is due to the relationships existing among the keys should yield only *one* set of intercorrelations. Moreover, the method used should have a logical basis. The fact that the intercorrelations of the "yes" weights, and those of the "no" weights approximated the intercorrelations of total trait scores may be an *a posteriori* reason for supposing that this method is legitimate, though Lorge advances no argument for its use.

The purpose of the present article is to develop a formula for estimating the amount of correlation due to the relationships existing among the keys, and to compare this formula with other methods, both as to logical bases and as to results.

---

[1] Lorge, Irving: "Personality traits by fiat. I. The analysis of the total trait scores and keys of the Bernreuter Personality Inventory." *J. Educational Psychol.*, Vol. XXVI, April, 1935, pp. 273–278.

II. DERIVATION OF THE FORMULA

Let $X_1, X_2, \ldots, X_n$ be the scores obtained by applying the n keys in turn to the testees' patterns of responses. Let

$$\Gamma_1, \Gamma_2, \ldots, \Gamma_m$$

be the scores on the m items, each to be 1 or 0, accordingly as the response is "yes" or "no" (*or* like or dislike). Also let

$$\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1m}$$

be the differential weights assigned the items by Key 1, and likewise for the other keys, letting the first of the two subscripts designate the key and the second, the item number.

Then the score obtained by applying Key 1 to the responses will be

$$X_1 = \alpha_{11}\Gamma_1 + \alpha_{12}\Gamma_2 + \cdots + \alpha_{1m}\Gamma_m \tag{1}$$

Passing to the deviation scores,

$$x_1 = \alpha_{11}\gamma_1 + \alpha_{12}\gamma_2 + \cdots + \alpha_{1m}\gamma_m \tag{2}$$

$$\Sigma x_1 x_2 = \alpha_{11}\alpha_{21}\Sigma\gamma_1^2 + \alpha_{12}\alpha_{22}\Sigma\gamma_2^2 + \cdots + \alpha_{1m}\alpha_{2m}\Sigma\gamma_m^2 + \text{cross-products} \tag{3}$$

$$\Sigma x_1^2 = \alpha_{11}^2\Sigma\gamma_1^2 + \alpha_{12}^2\Sigma\gamma_2^2 + \cdots + \alpha_{1m}^2\Sigma_m\alpha^2 + \text{cross-products} \tag{3}$$

From (3) we may now write the formula for the correlation between scores obtained by applying Keys 1 and 2

$$r_{x_1x_2} = \frac{\alpha_{11}\alpha_{21}\Sigma\gamma_1^2 + \alpha_{12}\alpha_{22}\Sigma\gamma_2^2 + \cdots + \text{cross-products}}{\sqrt{\alpha_{11}^2\Sigma\gamma_1^2 + \alpha_{12}^2\Sigma\gamma_2^2 + \cdots + \alpha_{1m}^2\Sigma\gamma_m^2 + \text{cross-products}}} \cdot \frac{\alpha_{1m}\alpha_{2m}\Sigma\gamma_m^2 + \text{cross-products}}{\sqrt{\alpha_{21}^2\Sigma\gamma_1^2 + \alpha_{22}^2\Sigma\gamma_2^2 + \cdots + \alpha_{2m}^2\Sigma\gamma_m^2 + \text{cross-products}}} \tag{4}$$

This is the complete formula, and must give the exact correlation. However, in this form it is cumbersome to use, and a simplified formula may give good results.

If we assume that the items are uncorrelated,

$$r_{\Gamma_1\Gamma_2} = \frac{\Sigma\gamma_1\gamma_2}{\sqrt{\Sigma\gamma_1^2}\sqrt{\Sigma\gamma_2^2}} = 0$$

whence

$$\Sigma\gamma_1\gamma_2 = 0 \tag{5}$$

and all the cross-products will disappear, leaving the fomula much simplified. Even though the intercorrelations of items are not zero, they are small. Since some of the "yes" responses, for which $\Gamma = 1$, are favorable, and some are unfavorable, to the trait for which the key is made, some of the intercorrelations of items will be positive and some will be negative. This fact causes the sum of the cross-products to approach zero still more closely, and proves the reasonableness of the above assumption.

Applying the gross-score formula[1] to the item,

$$N^2\sigma^2 = N\Sigma\Gamma^2 - (\Sigma\Gamma)^2$$

but, since Γ is either 1 or 0,

$$\Sigma\Gamma = \Sigma\Gamma^2 = \text{number of persons passing the item,}$$

hence,

$$N^2\sigma^2 = N\Sigma\Gamma - (\Sigma\Gamma)^2$$

and

$$\sigma^2 = p - p^2 = pq \tag{6}$$

where p is the proportion of persons passing the item and q is the proportion failing it.

$$\Sigma\gamma_1^2 = N\sigma_1^2 = N_{p_1q_1} \tag{7}$$

If we drop the cross-products from (4), make substitutions from (7), and cancel the N's, we may write the formula for correlation

$$\gamma_{x_1x_2} = \frac{\alpha_{11}\alpha_{21}p_1q_1 + \alpha_{12}\alpha_{22}p_2q_2 + \cdots +}{\sqrt{\alpha_{11}^2p_1q_1 + \alpha_{12}^2p_2q_2 + \cdots + \alpha_{1m}^2p_mq_m}}$$
$$\frac{\alpha_{1m}\alpha_{2m}p_mq_m}{\sqrt{\alpha_{21}^2p_1q_1 + \alpha_{22}^2p_2q_2 + \cdots + \alpha_{2m}^2p_mq_m}} \tag{8}$$

Again we write

$$\sigma^2 = p - p^2$$

Transposing, and adding $\frac{1}{4}$ to both sides of the equation,

$$\sigma^2 + p^2 - p + \tfrac{1}{4} = \tfrac{1}{4}$$

whence

$$\sigma^2 + (p - \tfrac{1}{2})^2 = (\tfrac{1}{2})^2 \tag{9}$$

---

[1] Toops, Herbert A.: "Computing intercorrelations of tests on the adding machine." *J. Applied Psychol.*, Vol. VI, 1922, pp. 172–184.

This equation shows that, as p passes from 0 through 0.5 to 1.0, σ describes a semicircle with a radius 0.5 and with its center at (0, 0.5). From this we know that σ does not drop away rapidly from 0.5 until p approaches the limits, 0 and 1. Consequently the value of σ does not vary greatly within the middle range.

Hence, we can still further simplify our problem by assuming

$$\sigma_1 = \sigma_2 = \cdots = \sigma_m \tag{10}$$

In which case, from (6)

$$p_1q_1 = p_2q_2 = \cdots = p_mq_m$$

When we use this assumption, all of the pq factors will become unity when the numerator and the denominator are divided through by p_1q_1 and the correlation between two scores may be written

$$r_{x_1x_2} = \frac{\alpha_{11}\alpha_{21} + \alpha_{12}\alpha_{22} + \cdots + \alpha_{1m}\alpha_{2m}}{\sqrt{\alpha_{11}^2 + \alpha_{12}^2 + \cdots + \alpha_{1m}^2}\sqrt{\alpha_{21}^2 + \alpha_{22}^2 + \cdots + \alpha_{2m}^2}} \tag{11}$$

This formula is an algebraic statement of the rule given by Hull[1] and attributed to him by Poisson. Hull's problem was similar to our own. The rule is that for computing the theoretical correlation between two tests with common aptitude determiners. His determiners were uncorrelated, and we have assumed our items to be uncorrelated. We have had to assume in addition that the σ of the item scores were equal. Although Hull seemingly made no such assumption, this condition no doubt was closely approximated when he threw five thousand dice for each of his determiners, for the σ of any series of five thousand dice throws could not be far from the expected value of 1.708.

If this formula is correct, how was Lorge able to secure good results by correlating "yes" weights and "no" weights? It seemingly was due to a combination of circumstances. Upon examination it will be seen that the "yes" weight and the "no" weight for a given item on the Bernreuter keys are approximately equal in magnitude and are opposite in sign. If we let W represent the "yes" weight and W′ the "no" weight,

$$W = -W' \text{ approximately}$$

[1] Hull, C. L.: *Aptitude Testing.* World Book Company, 1928, pp. 230ff.

Since in our derivation we arbitrarily gave the "no" response a weight of zero,

$$\alpha = W - W' \tag{12}$$

$$\alpha = 2W \text{ approximately} = -2W' \text{ approximately} \tag{13}$$

By substituting from (13) in (11), the correlation of *scores* may be written

$$\gamma_{X_1X_2} = \frac{\Sigma W_1 W_2}{\sqrt{\Sigma W_1^2}\sqrt{\Sigma W_2^2}} \text{ approximately} \tag{14}$$

and likewise for the "no" weights.

The formula for the correlation of *weights* may be written

$$\gamma_{W_1W_2} = \frac{\Sigma W_1 W_2 - N M_1 M_2}{\sqrt{\Sigma W_1^2 - N M_1^2}\sqrt{\Sigma W_2^2 - N M_2^2}} \tag{15}$$

Since the means M_1, M_2, etc., are not far from zero,[1] (15) is an approximation of (14), and therefore of the correct Formula (11). This explains how the incorrect formula gave the results that it did.

The divergence between the results of (15) and (11) is introduced largely by the NM_iM_j term in the numerator and in the denominator radicands of (15), for (14) is a very close approximation of (11), as may be seen from Table I.

In this table, the results obtained by Formula (14) show an average discrepancy of only .0115, whereas those obtained by Formula (15) show an average discrepancy of .0361.

TABLE I.—COMPARISONS OF RESULTS OBTAINED BY FORMULAS

	Formula (11)	"Yes" weights formula (14)	"No" weights formula (14)	"Yes" weights formula (15)	"No" weights formula (15)
r_{12}	−.301	−.300	−.298	−.265	−.261
r_{13}	.856	.850	.830	.804	.800
r_{14}	−.737	−.702	−.730	−.744	−.714
r_{23}	−.157	−.150	−.151	−.082	−.092
r_{24}	.450	.442	.447	.434	.428
r_{34}	−.526	−.491	−.527	−.531	−.489

III. CONSTRUCTION AND CORRELATION OF CHANCE SCORES

In order further to investigate the Bernreuter Personality Inventory, one hundred patterns of responses were made by throwing dice.

[1] Lorge, Irving: *Op. cit.* Table II, p. 274.

A number of dice were thrown at a time and were lined up against the side of the box before the cover was removed in order to give a definite sequence for reading them. The first die determined the response for the first item, the second die for the second item, etc. If the die showed one, three, or five pips, the item was considered as marked "yes," but if the die showed two, four, or six pips, the item was considered as answered "no." The "yes" weights were punched into one end of a Hollerith card and the "no" weights into the other end. The card for an item answered "no" was turned end for end, thus bringing the no weights in line to add in the same counters on the tabulator as "yes" weights. Since there were just one hundred twenty-five cards in all, this arrangement greatly lessened the danger of counting both weights for one item, and compensating by failing to count either of the weights for another item. Following Bernreuter's example,[1] we disregarded the "?" answers and made all answers either "yes" or "no."

In doing this, the odd-numbered items were at all times kept separate from the even-numbered items. The odd-numbered item cards were put through the tabulator first and the total printed. As the machine struck the total, some of the counters cleared out, but others adding from the same fields on the card did not. When the even-numbered item cards had been put through, the summary punch punched the total scores and the evens scores. The odds scores were then punched by hand from the printed tape, since the machine had cleared these numbers from the windows by this time. In order further to speed up the process of making the chance scores, two identical decks of cards were used. Two assistants throwing the dice and arranging the cards for the next pattern of responses were nearly able to keep up with the tabulator. Since the evens scores and the total scores were punched with the summary punch, a check which made the odds-score total and the evens-score total agree with the total-score total was sufficient to insure accuracy of punching.

Thus there were obtained from each pattern of responses twelve scores—the total score, the odds score, and the evens score for each of the four keys. The intercorrelations of these scores were computed twice (as a check upon the accuracy of the computations) by means of the systematic method described by Toops.[2] The intercorrelations of the total score were computed separately from those of the parts scores. They are presented in Tables II and III.

---

[1] Bernreuter, R. G.: "Chance and personality inventory scores." *J. Educational Psychol.*, Vol. XXVI, April, 1935, pp. 279–283.

[2] Toops, Herbert A.: *Loc. cit.*

TABLE II.—INTERCORRELATIONS OF TOTAL CHANCE SCORES ON THE BERNREUTER PERSONALITY INVENTORY

	B1-N	*B2-S*	*B3-I*	*B4-D*
B1-N	1.000	−.190	.869	−.733
B2-S	−.190	1.000	−.087	.337
B3-I	.869	−.087	1.000	−.574
B4-D	−.733	.337	−.574	1.000
M	16.67	0.93	16.58	18.40
σ	29.25	24.99	20.57	23.48

TABLE III.—INTERCORRELATIONS OF PARTS CHANCE SCORES ON THE BERNREUTER PERSONALITY INVENTORY

		Odds				Evens			
		B1-N	*B2-S*	*B3-I*	*B4-D*	*B1-N*	*B2-S*	*B3-I*	*B4-D*
Odds	*B1-N*	1.000	−.258	.838	−.704	−.042	−.061	−.065	.022
	B2-S	−.258	1.000	−.164	.428	.149	−.030	.158	−.087
	B3-I	.838	−.164	1.000	−.436	−.039	−.080	−.064	−.047
	B4-D	−.704	.428	−.436	1.000	.082	−.090	.075	−.146
Evens	*B1-N*	−.042	.149	−.039	.082	1.000	−.245	.907	−.724
	B2-S	−.061	−.030	−.080	−.090	−.245	1.000	−.106	.372
	B3-I	−.065	.158	−.064	.075	.907	−.106	1.000	−.602
	B4-D	.022	−.087	−.047	−.146	−.724	.372	−.602	1.000
M		5.30	2.25	6.44	12.25	11.37	−1.32	10.14	6.15
σ		19.03	19.56	14.68	16.76	23.02	16.15	15.38	19.08

It will be noticed that wherever odds scores are correlated with evens scores, the correlations are easily explained as chance deviations from zero, with the possible exception of two. The intercorrelations of the evens scores agree well with those of the odds scores, and both of them agree well with the intercorrelations of the total scores. The means of the total scores are somewhat lower than those reported by Bernreuter,[1] but the standard deviations agree fairly well with his.

IV. COMPARISON OF RESULTS

Table IV brings together these various results. In the first column are given the weighted means of the intercorrelations computed from

[1] Bernreuter, R. G.: *Op. cit.* Table I, p. 280.

actual scores by Lorge, Bernreuter, and Flanagan and presented by Lorge.[1] The correlations were transformed into z's, and these weighted proportionally to N-3 in finding the mean z. The mean z's were then turned back into r's.[2] The computation of these means was greatly facilitated by the use of tables of z's prepared by Edgerton.[3]

TABLE IV.—INTERCORRELATIONS OF SCORES ON BERNREUTER PERSONALITY INVENTORY

	Mean of four correlations of actual scores (1)	Theoretical correlations by formula (11) (2)	Correlations of one hundred chance scores (3)	Mean of odds and evens correlations (4)	Correlation of "yes" weights[1] (5)	Correlation of "no" weights[1] (6)	Correlation of "?" weights[1] (7)
r_{12}	−.431	−.301	−.190	−.251	−.265	−.254	−.220
r_{13}	.960	.856	.869	.877	.804	.804	.257
r_{14}	−.812	−.737	−.733	−.714	−.744	−.696	−.176
r_{23}	−.339	−.157	−.087	−.135	−.082	−.092	−.057
r_{24}	.565	.450	.337	.401	.434	.428	.215
r_{34}	−.705	−.526	−.574	−.524	−.534	−.489	−.100

[1] From Lorge: *Loc. cit.*

In the second column are presented the theoretical correlations computed by Formula (11). In the third column are presented the correlations of the total chance scores from Table II. In the fourth column are given the means of the odds and evens correlations from Table III. In the fifth, sixth, and seventh columns are given the correlations of the "yes," "no," and "?" weights, as presented by Lorge.[4]

At a first glance, it would seem that some of the chance correlations in Column 3 deviate too far from the theoretical correlations in Column 2. An investigation by the z method shows that random samples of one hundred could vary as widely as the correlations in Column 3 do. Using Fisher's[5] technique, the critical ratios for the six correlations (that is, the difference between the z's divided by the standard error of

[1] Lorge, Irving: *Op. cit.* Table I, p. 274.

[2] Fisher, R. A.: *Statistical Method for Research Workers.* Edinburgh: Oliver and Boyd, 1930, Third Edition, pp. 170ff.

[3] Edgerton, H. A.: *Transformed Correlations.* Mimeographed, 9 pp.

[4] Lorge, Irving: *Loc. cit.*

[5] Fisher, R. A.: *Op. cit.*, p. 169.

the chance z) are found to be within the limits explainable by chance. They are presented in Table V.

For the two correlations where the critical ratio is large (namely, r_{12} and r_{24}) it was noticed that both the odds correlations and evens correlations from Table III gave a closer approximation of the theoretical correlation than the total-score correlation did. For that reason the two correlations were averaged for Column 4. It would have been much more desirable to make up two or three hundred more chance scores to see whether the correlations from the larger numbers would approach the theoretical correlations more closely.

TABLE V.—DEVIATION OF CHANCE CORRELATIONS FROM THEORETICAL CORRELATIONS IN TERMS OF THE STANDARD ERROR
(Using the z-technique)

CORRELATION	CRITICAL RATIO
r_{12}	1.16
r_{13}	−.50
r_{14}	.09
r_{23}	.70
r_{24}	1.32
r_{34}	−.68

By comparing Columns 5, 6 and 7 with Column 2, it can be seen that the method of correlating item weights does not give results good enough to warrant the use of this method in place of Formula (11).

A comparison of Columns 1 and 2 shows in a striking way that much of the correlation found between scores obtained by applying a number of the keys in turn to a given pattern of testee responses, on the Bernreuter Personality Inventory at least, is due to the relationships existing among the scores.

This correlation *may* exist in the general population, and the correlations among the criterion scores found on the standardizing population may be adduced in support of this view. But since such large correlations are built into the keys, let us be chary about accepting correlations of scores obtained by using these keys as evidence for it.

The differences between Columns 1 and 2 are due to the facts that the items are not uncorrelated, and that their σ's are not all equal, as we assumed in developing Formula (11).

V. SUMMARY

A formula was derived for estimating the correlation to be expected between scores obtained by scoring a given set of papers more than once.

It was explained how an incorrect formula had happened to give good results.

Attention was called again to the fact that correlations of successive scores obtained from a given set of test papers should not be offered as evidence of the existence of that correlation in the population at large.

OVER-COMPENSATION IN FINGER TIMING AS A FACTOR IN TYPEWRITING ERRORS

NEIL WARREN

University of Southern California

When it is intended that the two hands shall move simultaneously there is a consistent tendency for the hand not used in writing to precede the "preferred" hand. Metfessel and Warren,[1] whose results agreed with those of other studies, believed that the prior movements were caused by the tendency to over-compensate with the less-used hand. It has been shown, in addition, that when two fingers are moved together the less-used finger will consistently precede.[3] For example, the little finger of either hand will move before the forefinger of the same, or of the opposite hand. This, again, seems best xplained on the basis of over-compensation; of making greater allowance for the inferior member than is required by objective differences.

The interval by which one finger precedes the other in attempted simultaneous movement is ordinarily very short. In most bimanual activities its effect on the temporal relationship of the hands would be of little practical consequence. The accuracy of some such activities, however, may be effected by the tendency to over-compensate with the less-used member. The relative movements of coördinated digital processes are more likely to be affected by the tendency.

Typewriting Errors.—The tendency toward prior movement by the less-used finger in the various combinations may be of significance in such finger activities as typewriting, linotyping, and piano playing. Of such performances typewriting best lends itself to study since a record of the movements involved is produced by the activity itself. In typewriting errors those which might be a result of the tendency are transpositions, in which the error is a result of incorrect order of finger movements.

ANALYSIS OF TRANSPOSITIONS

In the study of this problem an analysis has been made of three hundred fifty-three transpositions, comprising 16.5 per cent of the letter errors in typewriting tests of sixty high-school pupils in classes in typing III and IV. These tests were made over a period of three weeks.

Arbitrary Classification.—For purposes of classification the fingers of the right hand have been numbered *R*-1, *R*-2, *R*-3, and *R*-4, beginning with the index finger. The digits of the left hand are similarly numbered *L*-1, *L*-2, *L*-3, and *L*-4. The fingers have been ranked arbitrarily in the order of their use from the least to the most in the following way: *L*-4, *R*-4, *L*-3, *R*-3, *L*-2, *R*-2, *L*-1, and *R*-1. Thus, in the usual system of "touch" typewriting, the letters *NA* would be typed by the fingers *R*-1 and *L*-4 in that order. If the finger movements were attempted simultaneously the results of the study reported by Warren[3] indicate that *L*-4 would move first and reverse the letters in about seventy per cent of the trials. Ordinarily, of course, it is intended that *N* should precede *A* by an appreciable interval of time so that transpositions need not occur so frequently.

It is significant, therefore, that as the speed of typewriting increases and the succeeding strokes approach more nearly to simultaneity, the proportion of errors which are a result of transpositions increases rapidly. In typing I the average number of strokes per minute was 61.6, and 2.1 per cent of the letter errors were reversals of finger order. The average number of strokes in typing III and IV was 207.5 and the percentage of transpositions was 16.5. That is, while the speed of typewriting increased 3.4 times, the proportion of transpositions to the total letter errors increased 7.8 times. At the same time the total number of errors decreased slightly in relation to the number of strokes made, being only .9 as great.

RESULTS OF ANALYSIS

The results of the tabulation of transpositions are given in Tables I and II. In Table I each finger is compared with those which receive less use. The fingers are arranged in each table in the order in which they exhibit the influence of the tendency toward over-compensation by preceding more-used fingers, and by following less-used ones. The fourth finger of the right hand is omitted because of insufficient data; it was involved only eight times.

Precedence of Less-used Fingers.—It will be noted in Table I that the least-used fingers (*L*-3 and *L*-4) are most likely to precede, rather than to follow more-used ones. *R*-2 and *R*-3, at the lower end of this list, are found in Table II to show the greatest tendency to follow, rather than precede, less-used fingers. In general, the indications are that the less used a finger is, the more likely it is to appear first in transposition errors, while the more-used fingers tend to appear

second. The four least-used fingers (*L*-4, *L*-3, *R*-4, and *R*-3) precede more-used fingers one hundred thirty-one times while they follow the same fingers seventy-two times. The four most-used fingers (*R*-1, *R*-2, *L*-1, and *L*-2) follow less-used fingers in one hundred seventy errors and precede them in one hundred thirty-two. In 61.4 per cent of the transpositions involving both hands, the left hand incorrectly

TABLE I.—RELATIONSHIP OF EACH FINGER TO MORE-USED FINGERS IN TYPEWRITING TRANSPOSITIONS

Finger	Precedes more-used fingers		Follows more-used fingers	
	Number	Per cent	Number	Per cent
L-5	42	77.8	12	22.2
L-4	58	67.4	28	32.6
L-1	33	57.9	24	42.1
L-2	34	50.0	34	50.0
R-3	30	50.0	30	50.0
R-2	12	48.0	13	52.0

TABLE II.—RELATIONSHIP OF EACH FINGER TO LESS-USED FINGERS IN TYPEWRITING TRANSPOSITIONS

Finger	Follows less-used fingers		Precedes less-used fingers	
	Number	Per cent	Number	Per cent
R-3	33	82.5	7	17.5
R-2	28	65.1	15	34.9
R-1	76	56.3	59	43.7
L-1	54	54.5	45	45.5
L-3	4	50.0	4	50.0
L-2	12	48.0	13	52.0

responded before the right. There were twice as many errors caused by a finger of the left hand preceding the corresponding finger of the right hand as caused by the right hand moving first.*

* It may be pointed out that in the "touch system" of type-writing the fingers rest upon a certain set of keys (*L*-4 on *A*, *L*-3 on *S*, *L*-2 on *D*, *L*-1 on *F*, *R*-1 on *J*, *R*-2 on *K*, and *R*-3 on *L*) and that one of these keys may be struck by the finger resting upon it while another finger in reaching for a key in another row. This would account for only 20.1 per cent of the errors. Moreover, of the errors not involving the "guide" keys, the percentages remain virtually the same as given in Tables I and II.

It is significant that 57.5 per cent of the transpositions involved the third and fourth fingers, while 29.7 per cent of the letter substitution errors in general were caused by these fingers. Pillé[2] reported that 17.8 per cent of the letter substitution errors involved the third and fourth fingers. Only the most frequent transpositions of her study are listed; of these, however, 79.0 per cent of one hundred thirty-eight errors are caused by prior movement of a less-used finger.

CONCLUSIONS

1. There is a consistent tendency for the less-used finger to precede a more-used one, when it is intended that the two shall move simultaneously.
2. The tendency is sufficiently definite to affect finely coördinated digital activities.
3. Analysis of typewriting errors indicates that the order of finger movements may be influenced so as to result in transpositions.
4. While all transpositions cannot be a result of the tendency to overcompensate, the tendency is a factor to be considered in methods of instruction in typewriting.

BIBLIOGRAPHY

1. Metfessel, M., and Warren, N. "Over-Compensation by the non-preferred hand in an action-current study of simultaneous movements of the fingers." *J. Exper. Psychol.* Vol. XVII, 1934, pp. 246–256.
2. Pillé, E. *An Investigation of First Semester Typewriting Errors.* MA. Thesis, University of Southern California, June, 1933.
3. Warren, N. "Over-compensation in time relationships of bilateral movements of the fingers." *Amer. J. Psychol.*, Vol. XXXXVII, pp. 580–596.

BISERIAL *r* *VERSUS* PEARSON *r* AS MEASURES OF TEST-ITEM VALIDITY

FRANCES SWINEFORD

Research Assistant, University of Chicago

One of the most troublesome questions which arises in the construction of tests is the selection of a method for validating the test items. Much attention has been given to the various available methods, but workers in the field fail to agree as to which is better, the Pearson correlation or the Biserial correlation for dichotomously scored items. The proponents of the former hold that the distribution of scores on an item which is scored "right" or "wrong" is discontinuous and cannot be considered normal in form. It is the purpose of this paper to present arguments in favor of the Biserial correlation in the case of items for a test which measures an ability or characteristic which is normally distributed throughout the population. The labor of the computation of the various measures of validity has been treated in an earlier discussion[1] and will not be taken into account here.

The theory that certain physical and mental characteristics are normally distributed has been generally accepted. A well-balanced test of any of these characteristics may be expected to give results whose distribution closely approximates the normal curve. If, however, the test contains a preponderance of easy or of difficult items, the distribution of scores will be negatively or positively skewed. By this reasoning it may be argued that the distribution of test scores, per se, does not define the distribution of the characteristic which the test measures. Without evidence to the contrary, a normal distribution may frequently be assumed, for "the normal curve is the best single approximation to most educational data."[2]

A single test item might itself be conceived as a test of the characteristic in question. Under the assumption that the characteristic is normally distributed, the scores on the single-item test may also be assumed to be normally distributed. The fact that the responses cannot be assigned to more than two categories need not alter such assumption. In such case the use of the Biserial *r* is justified.

---

[1] Swineford, Frances: "Validity of Test Items." *Journal of Educational Psychology*, Vol. XXVII, January, 1936, pp. 68–78.

[2] Holzinger, Karl J.: *Statistical Methods for Students in Education*. Boston: Ginn and Company, 1928, p. 221.

The important advantage of the Biserial r over the Pearson r is the fact that the former is not affected by extreme values of the difficulty of the item. The Pearson r is considerably lowered for items whose difficulty rating lies outside the range of twenty to eighty per cent. For purposes of illustration, both the Biserial r and the Pearson r were computed for one hundred ninety-one mental-test items ranging in difficulty from five to ninety-five per cent. Their respective coefficients of correlation with balance (the smaller of p and q) are .0055— and .2714. The correlation between the values determined by the two methods is .9371, which becomes .9655— when seventeen items whose balance is less than ten per cent are eliminated. The mean Biserial r for the one hundred ninety-one items is .451 and the mean Pearson r, .327. For the one hundred seventy-four items ranging in difficulty from ten to ninety per cent these means are .455— and .338, respectively. The corresponding means of the seventeen items whose balance is between five and ten per cent are thus .411 and .211.

The mathematical relationship between the two methods is easily derived and may be written as follows:

$$\frac{\text{Biserial } r}{\text{Pearson } r} = \frac{\sqrt{pq}}{z} \tag{1}$$

For purposes of calculation it is suggested that the Pearson formula be used written in the following form:

$$\text{Biserial } r = \frac{p(\bar{Y}_2 - M_y)}{\sigma_y z} \tag{2}$$

in which the variable elements are p, z, and $\bar{Y}_2$; M_y and σ_y being constant for all the items of the test. A table of $p/\sigma_y z$ can readily be made with the aid of existing tables of z for successive values of p.

ADDITIONAL DATA ON BINOCULAR IMBALANCE AND READING

BRANT CLARK

Psychology Laboratory, University of Southern California

INTRODUCTION

In a previous paper[1] data were presented to show that during a short reading period the eye movements of subjects having high exophoria (a tendency of the eyes to turn out) did not differ reliably from a control group in regard to the average number of fixations per line, the average number of regressions per line, etc. However, the exophores did show greater over-convergence movements as the eyes moved from the end of one line to the beginning of the next although the difference was not highly reliable. In connection with another study six subjects who exhibited very high exophoria (fifteen to twenty-four diopters at the reading distance) were made available and were photographed to determine the nature of their eye movement habits. These subjects, in contrast with the previous group, were suffering ocular discomfort as a result of the faulty condition of their eyes.

PROCEDURE AND RESULTS

The experimental procedure was the same as that used in the previous experiment. The subjects read the same paragraphs while their eye movements were photographed. The ages of the subjects of this group ranged between sixteen and thirty-five years. This variation in age and familiarity with the material did not permit a comparison of this group with the original control group for the usual reading measures, so the photographic records were examined only to determine the magnitude and time of the divergence movements which compensated for the over-convergence of the eyes in moving from the end of one line to the beginning of the next. The average extent of the corrective movements for this experimental group was found to be 102 ± 34.8 minutes as compared to 30.0 ± 9.5 minutes for the control group. D/σ diff. was found to be 5.5. Divergence movements as large as two hundred four minutes (the angular equivalent of approxi-

[1] Clark, B.: "The effect of binocular imbalance on the behavior of the eyes during reading." *Journal of Educational Psychology*, Vol. XXVI, 1935, pp. 530–538.

mately two cm. or three words on the page) occurred, and movements as large as one hundred fifty minutes were very common. Corrective divergence movements occurred after ninety-eight per cent of the return sweeps.

The average time required to complete these divergence movements was found to be 3.33 ± 0.71 twenty-fifth seconds while the time for the control group was 2.34 ± 0.55 twenty-fifth seconds. (D/σ diff. 3.3.) In other words, the control group required, on the average, about $\frac{1}{25}$ second less time to complete the divergence movements at the beginnings of the lines.

The data also tend to indicate that there are other ocular factors which may be important in producing this over-convergence because the extent of the binocular imbalance was far from perfectly related to the extent of the over-convergence movements, *e.g.*, the subject with the largest binocular imbalance (twenty-four diopters) made corrective movements which averaged sixty-eight minutes while the subject with the smallest binocular imbalance (fifteen diopters) made corrective movements which averaged one hundred forty minutes. On the other hand, this group as a whole made larger corrective movements than did the previous experimental and control groups which had smaller binocular imbalances. There were, however, too few subjects in this group to draw definite conclusions concerning this.

These data lend support to the findings of the previous study in that individuals with high binocular imbalances made larger "fusion" movements than the subjects of a control group and that any additional ocular fatigue resulting from these movements is probably not a fatigue of the extra-ocular muscles for the individual "fusion" movements were much slower than and approximately the same size as the interfixation movements. It is an established fact, however, that "fusion" movements of this sort are very fatiguing, and in an extended reading period they probably cause enough fatigue to be of considerable importance in a remedial reading program.

SUMMARY

A comparison of the eye movement records of subjects having high exophoria, which caused ocular discomfort, with the records of a control group under the limited reading period of this experiment leads to the following conclusions:

1. The exophoric group made significantly greater divergence movements at the beginning of the lines.

2. The exophoric group required a significantly greater amount of time to complete these divergence movements.

3. The large divergence movements of the subjects with high binocular imbalances probably cause enough fatigue to be important in a remedial reading program.

4. Other ocular factors than the high exophoria may also be important in causing these large divergence movements.

BOOK REVIEWS

FLOYD H. ALLPORT, LYNNETTE WALKER, and ELEANOR LATHERS. *Written Composition and Characteristics of Personality.* New York: Archives of Psychology, No. 173, 1934, pp. 82.

This monograph reports a study of students' themes from the point of view that instruction should not be confined to bringing the student up to mechanical standards of good rhetoric and grammar, but should also attempt to develop his ability to express his personality, *i.e.*, to do creative writing. The experiments involve the development of an index of identifiability for each of seventy students, based on the number of cases out of a possible eight in which readers are able to match themes according to authorship. In any matching, themes from all seventy students are not read together, but are taken in groups of five. Thus in the first matching, two themes from each of five students are presented to the reader, who attempts to pair them before taking another group. In the second matching, the reader tries to match five new themes (third writing) with five pairs from the first two writings, after the first matching has been corrected. In the third matching the reader tries to identify each new theme with a particular student's three previous themes, etc. A student is not grouped with the same four other students in succeeding matchings, themes on each of the three types of topics are not kept together, and practise and fatigue of the readers are equalized by having some of a student's themes read early and some late in the experiment. In fact, all such factors which might reasonably be suspected as sources of error are systematically varied by an elaborate program which might well serve as a laudable example of experimental series control.

In making matchings, the reader makes a progressive chart of hypothetical traits of the writer recognizable in his writing, which may serve to characterize him and to identify his future themes. One weakness of the experiment is the identification of themes on the basis of errors rather than entirely on personality characteristics. One wonders if the readers were properly instructed to avoid this pitfall. It occurs to the reviewer that it could have been avoided by having the themes corrected to remove mechanical errors before presenting them to the readers. For the purpose of correlating individuality with students' grades, themes could have been graded after having been corrected.

While we agree with the authors' justification (p. 16) of the use of only two readers, the low r between them, together with the considerable validity of each above chance, indicates the desirability of more readers. Also, if a judgment of a theme is partly characteristic of the personality of the reader as well as of the writer, or if traits of the reader partly blind him to certain traits of the writer, then a larger number of readers is indicated.

To test the hypothesis (p. 30) that the school discourages both too much and too little individuality, the authors might have correlated the grades with deviations of the identification index from the mean without regard to sign. Of course, the distribution of deviations would not be normal, and the Pearsonian coefficient could not be compared with coefficients derived from normal distributions, but this would not be necessary for the present purpose. From scattergrams and calculations made by the reviewer, he judges the relation too slight to support the authors' hypothesis. The authors are to be commended for presenting their raw data in full, a practise which should be more popular with experimenters.

The formulae on page 32 cannot give the values above fifty which are reported unless in the first formula the first minus sign is changed to plus, in which case the second formula is unnecessary.

The authors present an interesting defense of the theory of personality as a possession of the individual, against the theory of personality as a social relation between the individual and the observer. Accordingly they hold that the correct value of the identification index for an individual is not the average of the two values obtained by the two observers but the higher value, on the ground that it is *prima facie* obtained by the one who is more accurate in observing that individual. Then the most accurate score would be the highest from a large number of observers. Now with an infinite number of observers, each subject would receive the maximum possible score from some observer or other, hence the true scores of all subjects would be equal. Thus the authors' theory of the identification index reduces to absurdity. (If it follows necessarily from their theory of personality, then the latter is false to the extent that it demands their theory of the index.) The alternative concept of the index (that the true value is not the highest value obtained but the average from the various observers) seems to us to be demanded by the index itself. The authors say that upon this alternative theory the index would reflect not what the subject's "individuality

is, but . . . *how much* individuality he" has. Doesn't a quantitative measure on a unidimensional scale always imply not what but how much?

The results of the experiments show that each reader matches the themes correctly in fifty per cent of the cases, which is 2.5 times chance expectation under the present conditions. The basis of identification of the themes is not entirely known but is partly the personality of the student as revealed in his writing, partly the mechanics of his writing, plus (apparently) the reader's reactions to characteristics of the themes. The question as to whether the methods of teaching influence the identifiability of themes is not answered, probably because of the small number of data, as there is some indication that reading themes in class and having students discuss them encourages individuality more than does formal classroom procedure. No relation is found between individuality and: Grades on themes, improvement in grades, length of training, nor intelligence. About thirty pages are given to case studies and psychological analysis of personality as revealed in themes and its relation to faults of writing. This portion of the study should be of vital interest to teachers of composition. It indicates that a consistent sketch could often be made of one's personality as revealed in his writing, and that some errors are not isolated occurrences to be corrected by drill but may spring from underlying personality traits. C. R. GARVEY.

Carnegie Institute of Technology.

MADISON BENTLEY AND E. V. COWDRY. *The Problem of Mental Disorder*. New York: McGraw-Hill Book Company, 1935, pp. 388.

Can a fundamental knowledge of nature and of man be used in a broader and more effective way to improve our understanding of mental disorders and presently lead to more effective control? That is the question Dr. Bentley, chairman of a committee on psychiatric investigations carried on under the auspices of the National Research Council, sets up for the committee in the preface to a discussion of the findings. The manner of answering the question, so far as one can gather from the contents of the book published by the committee, is to survey current points of view in the field of psychiatry and to search through the many supporting sciences for their possible contributions. The bulk of the book is concerned with this latter purpose,

and a description of present contributions of twenty supporting sciences discussed by that number of authors. Some of the fields considered are: Cerebral anatomy and physiology, electrophysiology of the brain, neuropathology, biology and genetics, endocrinology, nutrition, experimental psychology, anthropology, and education. Five points of view in the fields of psychiatry are presented: Clinical psychiatry by Dr. C. Macfie Campbell, medical psychiatry by Dr. A. Meyerson, the neurologist's point of view by Dr. I. S. Wechsler, the psychobiological point of view by Dr. Adolf Meyer, and the psychoanalyst's point of view by Dr. Lawrence S. Kubie. All students of the nature of man will be interested in this serious attempt to bring together the fields of fundamental knowledge with the field of curative art. This book should be of particular importance, of course, to individuals dealing with human problems in a professional manner. To such individuals the content of this book is likely to give a more adequate point of orientation for evaluating all investigations and therapeutic methods that are now emerging. H. MELTZER.

Psychological Service Center, St. Louis.

C. S. BLUEMEL. *Stammering and Allied Disorders.* New York: The Macmillan Company, 1935, pp. 182.

Stammering, according to Dr. Bluemel's latest conception, is a partial or intermittent inhibition of conditioned reflexes. Hence in his latest book about stammering more than one-half of the content is devoted to a discussion of the topic of conditioned reflexes, conditioned inhibition, and language and speech considered from the point of view of reflexes and inhibitions. In this book a good deal of space is also devoted to a re-interpretation of the author's own earlier auditory amnesia explanation, Dr. Swift's visual imagery explanation and even Dr. Travis's gradient explanation of stammering in terms of his present conception of conditioned inhibition. Of his own earlier conception he says that it is not materially altered by the introduction of the term "inhibition" but instead is clarified by the change in terminology. Certainly most American students of psychology will be glad of this change in terminology.

In a relatively short chapter on the treatment of stammering Dr. Bluemel suggests that therapy would be simplified if the mechanics of speech were disregarded and if attention were directed to the production of the finished word and sentence, regarded as conditioned

reflexes. Medical, psychological and social aspects of treatment are also briefly discussed in this chapter. If the book does nothing more than give some of the speech teachers who are not too well trained in psychological thinking a license to familiarize themselves with psychological concepts and not confuse themselves with worries concerning auditory or visual amnesia explanations it will have rendered a useful service.

H. MELTZER.

Psychological Service Center, St. Louis.